UNITED STATES AIR FORCE

SUMMER RESEARCH PROGRAM -- 1995

SUMMER FACULTY RESEARCH PROGRAM FINAL REPORTS

VOLUME 2A

ARMSTRONG LABORATORY

RESEARCH & DEVELOPMENT LABORATORIES

5800 Uplander Way

Culver City, CA 90230-6608

Program Director, RDL
Gary Moore

Program Manager, AFOSR
Major David Hart

Program Manager, RDL
Scott Licoscos

Program Administrator, RDL
Gwendolyn Smith

Program Administrator, RDL
Johnetta Thompson

Submitted to:

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH

Bolling Air Force Base

Washington, D.C.

December 1995

REPORT DOCUMENTATION PAGE			Form Approved
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time and maintaining the data needed, and completing and reviewing the collection of information. Send comments rega information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for infor 1204, Arlington, VA 22202-4302, and to the Office of management and Budget, Paperwork Reduction Project (0704-0188			AFRL-SR-BL-TR-98- 0822
1. AGENCY USE ONLY (Leave Blank)	2. REPORT DATE December, 1995	3. REPORT TY Final	
4. TITLE AND SUBTITLE USAF Summer Research Program - 1995 Summer Faculty Research Program Final Reports, Volume 2A, Armstrong Laboratory			5. FUNDING NUMBERS
6. AUTHORS Gary Moore			
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Research and Development Labs, Culver City, CA			8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) AFOSR/NI 4040 Fairfax Dr, Suite 500 Arlington, VA 22203-1613			10. SPONSORING/MONITORING AGENCY REPORT NUMBER
11. SUPPLEMENTARY NOTES Contract Number: F49620-93-C-0063			
12a. DISTRIBUTION AVAILABILITY STATEMENT Approved for Public Release			12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

The United States Air Force Summer Faculty Research Program (USAF- SFRP) is designed to introduce university, college, and technical institute faculty members to Air Force research. This is accomplished by the faculty members being selected on a nationally advertised competitive basis during the summer intersession period to perform research at Air Force Research Laboratory Technical Directorates and Air Force Air Logistics Centers. Each participant provided a report of their research, and these reports are consolidated into this annual report.

14. SUBJECT TERMS			15. NUMBER OF PAGES
AIR FORCE RESEARCH, AIR FORCE, ENGINEERING, LABORATORIES, REPORTS, SUMMER, UNIVERSITIES			16. PRICE CODE
17. SECURITY CLASSIFICATION OF REPORT Unclassified	18. SECURITY CLASSIFICATION OF THIS PAGE Unclassified	19. SECURITY CLASSIFICATION OF ABSTRACT Unclassified	20. LIMITATION OF ABSTRACT UL

PREFACE

Reports in this volume are numbered consecutively beginning with number 1. Each report is paginated with the report number followed by consecutive page numbers, e.g., 1-1, 1-2, 1-3; 2-1, 2-2, 2-3.

Due to its length, Volume 2 is bound in two parts, 2A and 2B. Volume 2A contains #1-23, and Volume 2B contains reports #24-43. The Table of Contents for Volume 2 is included in both parts.

This document is one of a set of 16 volumes describing the 1995 AFOSR Summer Research Program. The following volumes comprise the set:

VOLUME	TITLE
1	Program Management Report
	Summer Faculty Research Program (SFRP) Reports
2A & 2B	Armstrong Laboratory
3A & 3B	Phillips Laboratory
4	Rome Laboratory
5A, 5B, & 5C	Wright Laboratory
6A & 6B	Arnold Engineering Development Center, Wilford Hall Medical Center and Air Logistics Centers
	Graduate Student Research Program (GSRP) Reports
7A & 7B	Armstrong Laboratory
8	Phillips Laboratory
9	Rome Laboratory
10A & 10B	Wright Laboratory
11	Arnold Engineering Development Center, Wilford Hall Medical Center and Air Logistics Centers
	High School Apprenticeship Program (HSAP) Reports
12A & 12B	Armstrong Laboratory
13	Phillips Laboratory
14	Rome Laboratory
15A&15B	Wright Laboratory
16	Arnold Engineering Development Center

SYSTEMATIC BIAS OF MEMORY FOR LOCATIONS IN TWO-DIMENSIONAL ARRAYS

Gary L. Allen
Associate Professor
Department of Psychology

University of South Carolina
Columbia, SC 29208

Final Report for:
Summer Faculty Research Program
Armstrong Laboratory

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, DC
and
Armstrong Laboratory

September 1995

SYSTEMATIC BIAS OF MEMORY FOR LOCATIONS
IN TWO-DIMENSIONAL ARRAYS

Gary L. Allen
Associate Professor
Department of Psychology
University of South Carolina

Abstract

Previous research indicated that memory for locations in two-dimensional graphic arrays is encoded as polar coordinates and is systematically biased such that locations are remembered toward the middle of quadrants that observers impose implicitly on the array. The results of the present experiment replicate these findings and further point to encoding time as a factor influencing the radial distance portion of this bias. These findings suggest the value of future experiments aimed at determining how other salient variables (such as a differentiated array space or multiple potential targets) impacts bias of this type.

SYSTEMATIC BIAS OF MEMORY FOR LOCATIONS IN TWO-DIMENSIONAL ARRAYS

Gary L. Allen

Introduction

Memory for specific locations within a two-dimensional array is an ability of practical and theoretical significance. In practical terms, this form of location memory is involved in the acquisition of information from maps and other graphic displays. This activity is common in some situations, such as using a map for wayfinding, and may be very important in some occupations, such as air traffic control. In theoretical terms, memory for spatial location has been described as a form of episodic memory that involves a combination of fine-grained and categorical information (Huttenlocher, Hedges, & Duncan, 1991). At present, there is strong evidence that categorical information, which is evidenced by the strong tendency for an observer to divide a visually presented display implicitly into regions, results in a systematic bias in estimating target location. This basic finding raises a number of interesting issues concerning categorical bias in particular and location memory in general. The potential importance of these issues for understanding human performance in a variety of contexts involving location memory constitutes a strong rationale for basic research aimed at these questions.

A reasonable point of departure for future research is the set of findings demonstrating that when research participants are shown a single dot within a circular field and then asked to estimate the dot's location on a comparable circular field in the absence of the original display, they tend to (a) compute their estimates on the basis of independent angular and radial distance components; (b) implicitly impose a horizontal and vertical axis on the display, thus dividing it into quadrants; and (c) bias their estimates towards the middle areas of quadrants (Huttenlocher et al., 1991). Their findings indicate that memory for spatial location in a two-dimensional array is determined by two distinct types of information: fine-grained and categorical. Fine-grained information results in distribution of estimates around the actual location; categorical information results in the previously mentioned displacement of estimates towards the middle of implicit quadrants.

This study was designed to test the replicability of Huttenlocher et al.'s (1991) findings using a computer-based procedure with different response demands and to test the hypothesis that fine-grained location information takes longer to

encode than does categorical information. Replicability, always an important issue, is particularly critical in this instance. Participants in all four of the studies reported by Huttenlocher et al. (1991) responded by mapping the remembered location from a presentation display onto a response display. The circle in the presentation display (presented either on paper or on computer screen) was the same size as that in the response display (presented either on paper or on a surface overlaying a digitizing pad), but participants looked straight ahead to observe the original display and then looked down to estimate stimulus location on the response display. This mapping of locations from one display onto another was eliminated in the current study. Instead, participants observed the location of the stimulus on a computer screen and then, after an intervening visual mask, indicated the location of the stimulus using a joystick-controlled position indicator. A priori it was not clear how the elimination of the response mapping component of the task would affect the categorical bias observed in the previous studies. Nevertheless, following the assumption that encoding fine-grained information requires more effort than encoding categorical information in this task and that eliminating the mapping component makes the task somewhat less demanding, it was considered possible that categorical bias would be diminished under the current procedure.

The other hypothesis examined in this study was that fine-grained and categorical memory for location are differentially influenced by limited encoding time. Although encoding time had not been examined by Huttenlocher et al. (1991) or by Nelson and Chaiklin (1980) before them, this prediction was based on the logic that finer-grained information takes longer to encode than does categorical (i.e., quadrant) information in this task. Accordingly, the estimated location of a briefly presented stimulus should be more influenced by categorical bias than should the estimated location of a stimulus presented for a longer period of time. The prediction of greater categorical bias is distinct from the quite sensible expectation that accuracy per se would diminish with limited encoding time.

Method

Participants. Data were collected from 40 Air Force enlistees who voluntarily participated during their sixth day of basic training at Lackland Air Force Base, Texas.

Task materials and equipment. The memory for location task consisted of 256 trials in which participants were instructed to indicate within a 15 cm diameter

circle the location where a 1.5 mm dot (referred to as a target) had been presented previously for a brief period of time within a circle of the same size. Each trial consisted of a auditory signal to alert the participant that a presentation display was forthcoming; a presentation display showing the target within the circle; a visual masking display consisting of eight 15 cm diameter half circles and an array of 1.5 mm dots whose locations appeared to move slightly, and a response display consisting only of a circle identical to the one shown in the presentation display. The duration of the signal and of the visual masking display (500 ms) were fixed. The duration of the response display was determined by response speed. After every group of eight trials, subjects were given feedback consisting of a symbol for each estimate within .5 cm of original target location. This feedback was designed to maintain participants' attention focused on the task without providing information that could influence response bias.

Two variables were manipulated: location of the target (128 possibilities: 32 angles by 4 radial distances along each angle) and duration of target presentation (50 ms versus 1 s). Locations were those used successfully by Huttenlocher et al. (1991), which included each combination of angle (0, 6, 19, 32.45, 58, 71, 84, 90, 96, 109, 122, 135, 148, 161, 174, 180, 186, 199, 212, 225, 238, 251, 264, 270, 276, 289, 302, 315, 328, 341, and 354 degrees from center) and radial distance (1.5, 3.0, 4.5, and 6.0 cm from center). Participants estimated location by aiming a target indicator with a joystick.

Procedure. The memory for location task was administered as part of a large battery of computerized tests that required approximately four yours of testing time. This specific task required approximately 20 minutes to complete. The task included instructions and four practice items. Test proctors were available to answer any questions during administration of the task.

Results

Separate 2 (presentation time) x 32 (angle) x 4 (radial distance) analyses of variance (ANOVAs) were performed as preliminary analyses on the signed angular error and the signed radial error of location estimates. The ANOVA performed on signed angular error yielded a single main effect for angle. Examination of means indicated that estimated angle of target location was more accurate for angles 0, 90, 180, and 270 than for others, thus corresponding to hypothesized implicit quadrant boundaries.

The ANOVA performed on signed radial distance error yielded main effect for presentation time, angle, and radial distance. Examination of means indicated that error of radial distance estimates was (a) greater after the 50 ms presentation time than after the 1 s presentation time; (b) greater for angles off of the hypothesized quadrant boundaries than for angles on those boundaries; and (c) greater for locations nearer the center of the circle than for distances nearer the boundary of the circle.

Discussion

The results from this experiment replicated that basic findings of Huttenlocher et al. (1991) in showing the systematic biasing of location memory in two-dimensional arrays. Specifically, accuracy in terms of angle of estimate varied as a function of the distance of the original target from implicit quadrant boundaries, and accuracy in terms of radial distance of estimate varied as a function of distance of the original target from the outer boundary of the response space. Current results extended previous findings by suggesting that radial distance error--but not angular error--is affected by encoding time. Limiting the exposure of the original target to 50 ms significantly increased the observer's bias in estimating distance from the center of the array.

These results support the theoretical interpretation that locations in two-dimensional arrays are encoded by two processes, one involving fine-grained, specific information and the other involving categorical, general information. The data are partially consistent with the interpretation that limiting encoding time leads to reliance exclusively on categorical, general information. The differential effect of encoding time on angular and radial accuracy deserves future examination. Investigating factors that drive either fine-grained or categorical encoding appears to be an important direction for future experiments.

One approach to examining the effects of fine-grained and categorical influences is to devise task circumstances that emphasize one type of information over the other. Categorical information could be emphasized by simply making the quadrants more distinct. Distinctiveness could be achieved by marking quadrants explicitly either by overlaying polar coordinate markers or by indicating the quadrants with different colors. The assumption is that making quadrants explicit would increase the tendency to encode locations categorically. At this point, it is difficult to predict whether quadrant distinctiveness would influence location

encoding, estimate production, or both. This question could be investigated by manipulating the presence or absence of overt quadrant markers during initial presentation of the display and during subsequent estimate production.

Proposing a manipulation that would increase relative emphasis on fine-grained location information is considerably more challenging. One possibility is to decrease the reliability of categorical information by randomly altering the functional response area during estimate production. Under such circumstances, research participants would view the normal target and display during encoding, but during estimate production, part of the circular display field--including part of the quadrant in which the target had been presented originally--would be shaded to indicate an out-of-bounds area. This manipulation would alter the shape of the functional response area and, more pointedly, the shape of the quadrant in which the target was located. It is hypothesized that the unreliability of quadrant shape would lead participants to rely more on fine-grained location information even on trials in which none of the response field was designated as out-of-bounds.

Another related problem that could be examined in future studies concerns the effect of multiple targets. Available evidence suggests that considerable information can be acquired from visual displays in a single glance, but the effects of memory load on the accuracy of location estimates are not well documented. These effects may depend on the spatial relations between the targets within the display. In general, it could be posited that multiple targets represent an increase in task demands in the same way that an intervening task does. To the extent that this position is true, it follows that multiple targets increase reliance on categorical encoding and estimate production. However, multiple targets within a common quadrant may have a very different influence on accuracy. If multiple targets within a single quadrant are presented without other targets in other quadrants, contrast effects may be observed in which the targets within the common category are estimated to be more distant from each other than is actually the case; consequently, the bias to locate targets toward the middle of implicit quadrants would be negated. However, if multiple targets are presented within a quadrant and other targets, especially distant ones, are presented, an assimilating effect may be observed in which fine-grained distinctions between targets within a common category are not made; consequently, the bias to locate targets toward the middle of implicit quadrants would be in evidence.

The multiple target tasks mentioned thus far do not necessarily require the differentiation of targets. In other words, the task requires the estimate of multiple target locations without designating a specific target-to-location relation. Of course, it is also possible to include target differentiation as a task factor. The obvious effect would be to increase memory load, thus leading to the general prediction of increased reliance on categorical information. Beyond this prediction, however, the considerations raised previously concerning the relative locations of the targets would apply.

References

Huttenlocher, J., Hedges, L. V., & Duncan, S. (1991). Categories and particulars: Prototype effects in estimating spatial location. *Psychological Review, 98,* 352-376.

Nelson, T. O., & Chaiklin, S. (1980). Immediate memory for spatial location. *Journal of Experimental Psychology: Human Learning and Memory, 6,* 529-545.

REDUCTIVE DEGRADATION AND SORPTION OF *cis*- AND *trans*-1,2-DICHLOROETHENE IN A METALLIC IRON/WATER SYSTEM

Richelle M. Allen-King, Assistant Professor and
Richard M. Halket, Graduate Student

Department of Geology
Washington State University
Pullman, WA 99164-2812

Final Report for:
Summer Faculty Research Program
Armstrong Laboratory

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, DC

and

Armstrong Laboratory

September 1995

REDUCTIVE DEGRADATION AND SORPTION OF *cis*- AND *trans*-1,2-DICHLOROETHENE IN A METALLIC IRON/WATER SYSTEM

Richelle M. Allen-King, Assistant Professor and
Richard M. Halket, Graduate Student
Department of Geology
Washington State University
Pullman, WA 99164-2812

ABSTRACT

Reductive transformation kinetic and sorption coefficients were determined for both *cis*- and *trans*-1,2-dichloroethene (DCE) in batch systems with zero-valent iron and water. Chloride was produced by the transformation reaction and chlorine mass balances for the batch systems were 80 to 85%. The transformation reaction was not first order in solution concentration or total system concentration for either of the two isomers. Measured reaction rate coefficients (λ_a) and orders (N_a) for the two compounds in experiments with initial concentrations of approximately 1850 nmol/ml were: 0.17 [nmol/hr]/[(nmol/ml)^{N_a}] (ln λ_a= -1.79) and 0.00023 [nmol/hr]/[(nmol/ml)^{N_a}] (ln λ_a= -8.37) with reaction orders 1.22 and 1.77 for *trans*-1,2-DCE and *cis*-1,2-DCE, respectively. Sorption equilibrium was apparently attained within 1.1 hr. The form of sorption could be adequately described by Freundlich-type isotherms for both compounds over the concentration range measured. The magnitude of sorption was greater for *trans*-1,2-DCE than for the more soluble *cis*-1,2-DCE. The distribution of organic products produced by the two isomers indicates some divergence in reaction pathways. While both compounds produced large proportions of ethene and ethane, transformation of *cis-1,2*-DCE resulted in significantly greater production of vinyl chloride than did *trans*-1,2-DCE.

REDUCTION KINETICS AND SORPTION OF *cis*- AND *trans*-1,2-DICHLOROETHENE IN A METALLIC IRON/WATER SYSTEM

Richelle M. Allen-King and Richard M. Halket

INTRODUCTION

Zero-valent iron can bring about the reduction of chlorinated solvents, such as perchloroethene and trichloroethene (PCE and TCE), dissolved in wastewater [Sweeny, 1981a; Sweeny, 1981b; Senzaki, 1991]. The reduction reaction is coupled to iron corrosion (oxidative dissolution) [Matheson and Tratnyek, 1994]. Because transformation rates are rapid relative to typical groundwater flow velocities, metallic iron is being tested in permeable *in situ* treatment barriers as a method for remediating contaminated groundwater [Gillham and O'Hannesin, 1994; Wilson, 1995].

In recent research, Burris et al. [1995] have demonstrated that sorption to nonreactive sites in iron/water systems is significant for PCE and TCE. For these two chemicals, the reduction reaction is: 1) first order in solution concentration when sorption to inert or nonreactive sites is accounted for; and 2) of a similar rate for the two compounds [Burris et al., 1995]. Low concentrations of dichloroethene (DCE) isomers and vinyl chloride (VC), as well as chloride ions, ethene, ethane, acetylene, methane, and C_3-C_5 alkanes have been detected as reaction products [Senzaki, 1991; Gillham and O'Hannesin, 1994; Burris et al., 1995; Campbell and Burris, 1995; Burris et al., in prep].

In order to accurately plan *in situ* remediation systems, a complete understanding of the reduction reaction pathways, rates and orders are needed. The fact that some of the potential products, such as VC, are themselves toxic makes this requisite knowledge for planning application of this technology. However, the degradation rates for intermediate reaction products, such as the DCE isomers, in the iron/water system have not previously been determined. They are reduced more slowly than TCE or PCE by microorganisms [Vogel et al.,

1987].

In this research, we focus on the reduction of *cis*- and *trans*-1,2-dichloroethene (*cis*-DCE and *trans*-DCE), to contribute to the overall goal of improving our understanding of the degradation scheme for chlorinated ethenes in the metallic iron/water system. The specific objectives of this study are to determine the reduction rates, reaction order and products, and sorption of *cis*-DCE and *trans*-DCE in a metallic iron/water system. Because the 1,2-DCE isomers are frequently identified as co-contaminants with PCE and/or TCE in groundwater and have been identified independently as groundwater contaminants [Westrick, 1990], the results will also be useful in evaluating the potential for remediation of 1,2-DCEs using zero-valent iron.

MATERIALS AND METHODS

Iron and chemicals

Fisher Scientific 40-mesh sized iron filings (lot number 946315) was used for this study. All iron was pretreated in 400 g batches by washing with 1 L of 1 N HCl for one-half hour then rinsing with 10 L of argon-sparged deionized water. After rinsing, the iron was dried overnight under a nitrogen atmosphere at 90 degrees C. Once dried the iron was stored in argon sparged glass jars to inhibit oxidation. Surface area determined by BET (nitrogen) analysis was 0.9 m^2/g for the untreated iron and 1.0 m^2/g for the treated iron [Burris et al., 1995]. Pyrite obtained from Wards Natural Science Establishment was used to pH buffer the batch system. Pretreatment of the pyrite consisted of powdering with a mortar and pestle and storing in argon sparged glass jars to inhibit oxidation.

Both *cis*- and *trans*-1,2-DCE were supplied by Aldrich Chemical Co. with purities of 97% and 98%, respectively. Methanol was Fisher HPLC grade and hexane was Fisher Optima grade. All water used in the experiments was treated with a Milli Q reagent water system.

Batch system

Two time-series experiments were conducted with each of the two DCE isomers at two different initial concentrations in batch systems. The batch systems were prepared by adding 10.00 g of pretreated iron and 0.20 g of powdered pyrite to argon-purged 5 ml nominal (10.2 ml observed) Wheaton serum vials. DCE-amended water was dispensed to the vials in an anaerobic glove box (5% H_2, balance N_2 atmosphere) to prevent oxidation during dispensing, and the vials were immediately sealed with teflon-lined rubber septa (Supelco). Iron- and pyrite-free controls were prepared to check for system losses attributable to mechanisms other than reduction. DCE-free controls were prepared with iron and pyrite to determine the chloride background concentration. After preparation, vials were shaken at 12 rpm at ambient laboratory temperature (22 to 25°C) until sampled. Four vials (two sample and two Fe-free control vials) were sacrificed at each sampling time for measurement of DCE solution concentrations and total DCE mass. Water samples (50 to 100 μL) were collected from the Fe-containing DCE and duplicate DCE-free control vials at each sampling time. Near the end of each time-series experiment, organic degradation products were determined by a headspace analysis, as described below.

Analytical

The aqueous DCE concentration was determined by transferring as much water as possible to a second 5 ml serum vial containing 3.00 ml hexane through a canula by air displacement. On average 2.40$\pm$0.16 ml of water remained in the Fe-containing vial after the transfer. The remaining DCE mass in the system, including mass sorbed, was extracted by adding 5 ml of methanol to the original batch vial and shaking for 15 minutes on a vortex mixer at high speed. Two milliliters of the resulting methanol/water mixture were diluted with 6.5 ml Milli-Q water and extracted with 4.5 ml hexane. The masses of total solution in the original vial before sampling, total methanol/water mixture, and methanol/water mixture sample transferred, were determined gravimetrically with an analytical balance (Metler). The hexane used for all extractions contained chloroform as an internal standard. Controls were treated the same way as samples with the exception that 6.5 ml of water was transferred to the receiving vial and methanol was added to the remaining water in the original sample vial. All samples were

analyzed on a Hewlett-Packard Model 5890 gas chromatograph equipped with flame ionization detector, Model 7673 autosampler, and a 60 m megabore DB-Wax (J&W Scientific) capillary column with a 1 μm coating. The GC operating parameters were: 2.9 ml/min column flow; 11.6 ml/min split flow; 3 μl sample injection; injector and detector temperatures were 200°C and 240°C, respectively; oven program was 50°C for 1 min, 3 °C/min to 98°C. The concentration of *cis*-DCE and *trans*-DCE were determined with a multilevel calibration curve obtained by measurement of standards with known concentration.

Near the end of each time-series experiment, a 1.00 ml headspace for analysis of volatile transformation products was created in a sample vial by displacing water with air. Following phase-equilibration at ambient temperature, a 100 μl subsample of the headspace was analyzed using the method of Campbell and Burris [1995].

Water samples for chloride (Cl^-) analysis were collected with a syringe prior to collecting the DCE water samples. Each sample was diluted with Milli-Q water to achieve a final concentration in the 0 to 20 mg/l calibration range. Samples were analyzed on a Dionex system equipped with a CDM-2 (conductivity) detector and an AS11 column. The chloride concentration was determined with a multilevel standard calibration curve.

RESULTS AND DISCUSSION

Time-series and sorption

Solution concentrations and total mass of the DCE isomers decreased throughout the course of an experiment in the Fe-containing samples (Fig. 1) while the concentrations were maintained near the initial value in the Fe-free control vials. On average 94% and 90% of the known mass added to the samples was recovered in the initial extractions for the *cis*- and *trans*-DCE isomers, respectively. These high relative recoveries (near 100%) demonstrate the success of the extraction technique in recovering the total mass added to the system. The lower recovery for the *trans*-DCE isomer is attributed to its faster reaction rate.

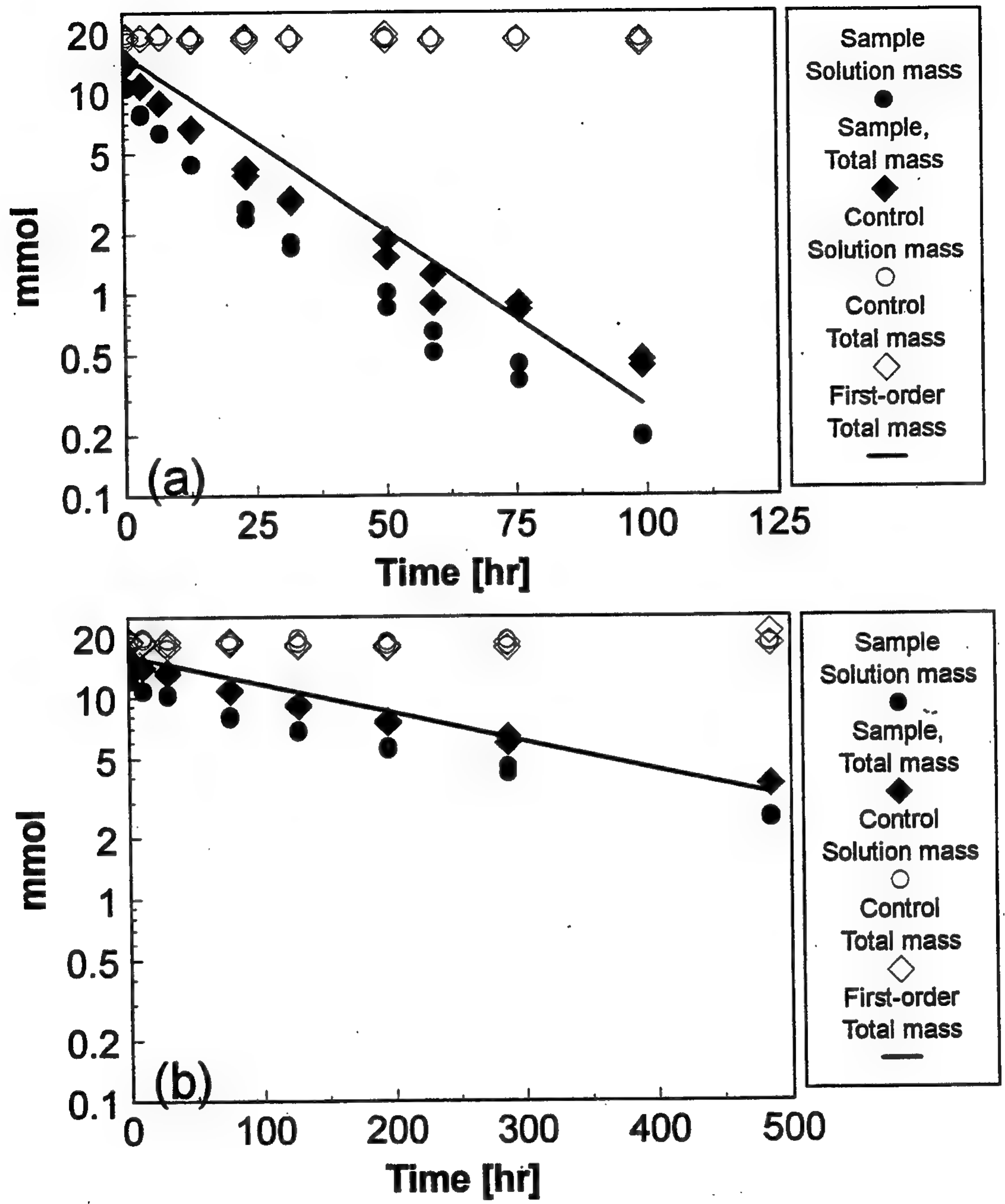


Fig. 1. Total and dissolved (a) *trans*-DCE and (b) *cis*-DCE masses over the experiment time course. *Trans*-DCE is transformed much more rapidly than *cis*-DCE. The difference between the total and dissolved masses is the sorbed mass.

Sorbed concentrations were determined for each sample from the difference between the total and solution DCE masses. Sorption isotherms for each of the two isomers are shown in Fig. 2. Sorption equilibrium was apparently achieved within 1.1 hr, as evidenced by the consistency between samples taken at 1.1 hr and at much longer times. Samples of the *trans*-DCE experiment collected after 0.5 hr were apparently not equilibrated with respect to sorption.

Sorption isotherms were fit to all equilibrium data with the Freundlich equation:

$$q_e = K_F C_a^{1/n} \tag{1}$$

where q_e is the sorbed concentration and C_a is the solution concentration. The coefficients determined are listed in Table 1. The greater magnitude of sorption for *trans*-DCE may be related to its lower solubility. The fact that the magnitude of sorption is greater for the more hydrophobic of the two solutes is consistent with previous observations. Sorption for both of the DCE isomers was much less than sorption of TCE and PCE observed in a previous study under similar conditions [Burris et al., 1995].

Table 1. Freundlich sorption coefficients (standard error) and number of observations (*m*) for *trans*- and *cis*-DCE.

compound	log K_F*	$1/n$	m
trans-DCE	0.471(0.032)	0.685(0.011)	40
cis-DCE	0.469(0.029)	0.658(0.013)	34

* Dimensions of K_F are $[\mathrm{nmol/g}]/[(\mathrm{nmol/ml})^{1/n}]$

Transformation rate coefficients and reaction orders

Rate coefficients (λ) and reaction orders (N) were determined for the following equations:

$$\frac{dM_T}{dt} = -\lambda_T M_T^{N_T} \tag{2}$$

and

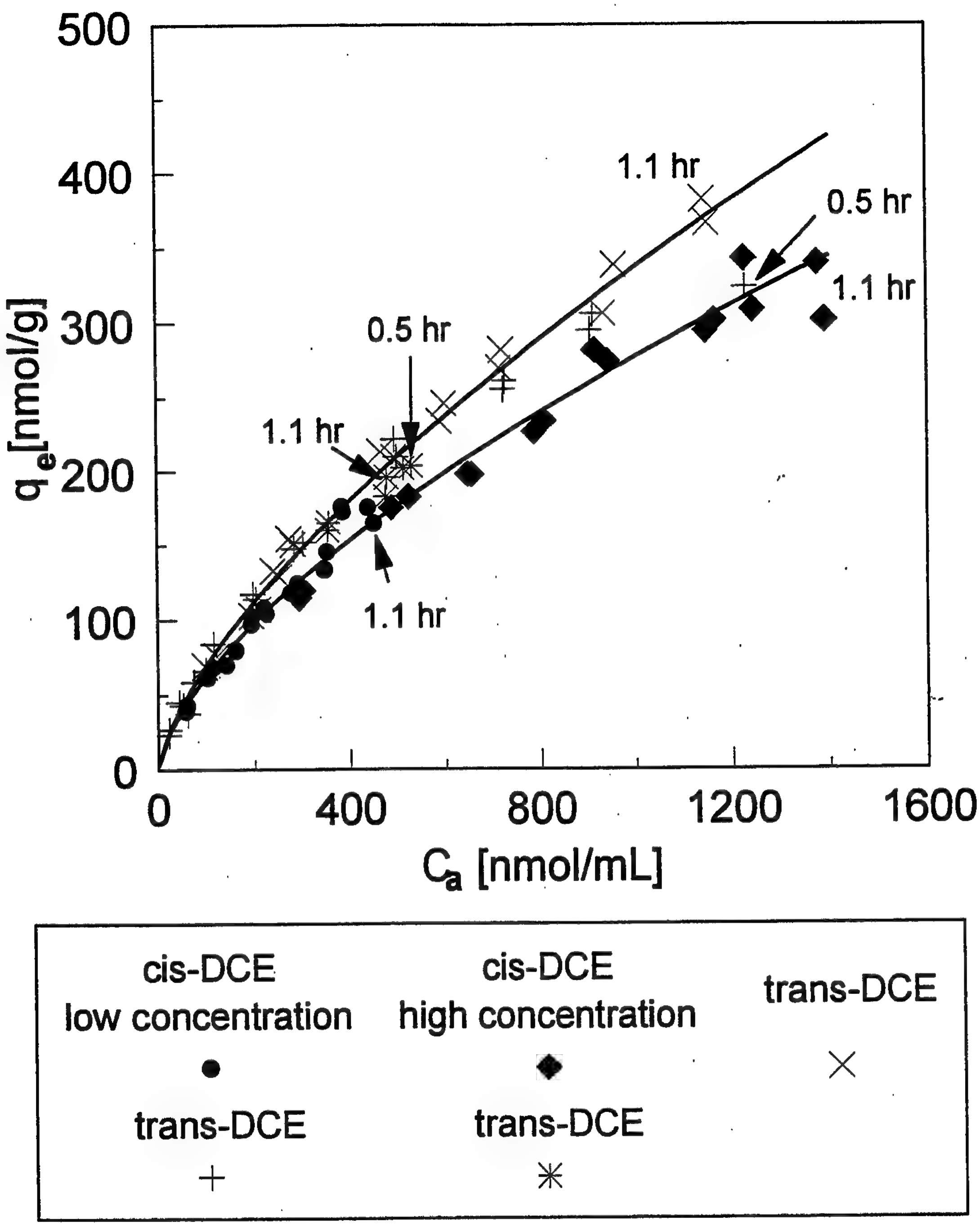


Fig. 2. Sorption isotherms. Sorbed concentration (q_e) is determined as the difference between the sorbed and total dissolved masses at each sampling time. Sorption equilibrium occurred within 1.1 hr.

$$\frac{dM_T}{dt} = -\lambda_a \, C_a^{N_a} \tag{3}$$

where M_T is the total DCE mass in the vial, superscripts T and a refer to coefficients determined based on total system mass and aqueous concentration, respectively. The equations were linearized by ln-transformation and the coefficients were determined by linear regression (Fig. 3 and 4). The results of replicate samples for each time were averaged.

The rate coefficients for *cis*-DCE were much lower than those for *trans*-DCE. Little difference between the N_a and N_t for each compound, compared to the differences observed for PCE for example [Burris et al., 1995], is attributable to the relatively low sorption which was observed. A visual comparison of the fit- (N_T) and first-order curves for *cis*-DCE is shown in Fig. 5. Clearly, the fit-order provides a better match of the measured data than the simpler first-order curve. The fit reaction orders for both DCE isomers were greater than unity for all cases, although N_a for *trans*-DCE was near unity. Previous research demonstrated that N_a was near unity for both TCE and PCE in similar experiments [Burris et al., 1995]. Therefore, our reaction order results are not consistent with the previous research, particularly for the *cis-DCE* isomer.

Table 2. Transformation rate coefficients (λ) and reaction orders (N) (standard error) for *trans*- and *cis*-DCE experiments.

Compound/ Experiment	ln λ_1	N_T	ln λ_T	N_a	ln λ_a
***trans*-DCE**	-3.22 (0.03)	1.41 (0.12)	-6.55 (0.37)	1.22 (0.11)	-1.79 (0.37)
***cis*-DCE, low**	-5.57 (0.02)	1.80 (0.25)	-11.85 (0.34)	1.64 (0.24)	-6.33 (0.35)
***cis*-DCE, high**	-5.74 (0.03)	1.89 (0.35)	-13.76 (0.34)	1.77 (0.33)	-8.37 (0.34)

The first-order rate coefficient (λ_1) is determined from the equation (after Schwarzenbach et al., 1993) $\lambda_1 = -(\ln M_T/M_{T,O})/t$, where $M_{T,O}$ is the initial total DCE mass (t = 0). Dimensions are: hr^{-1} for λ_1, $[nmol/hr]/[(nmol)^{NT}]$ for λ_T, and $[nmol/hr]/[(nmol/ml)^{Na}]$ for λ_a.

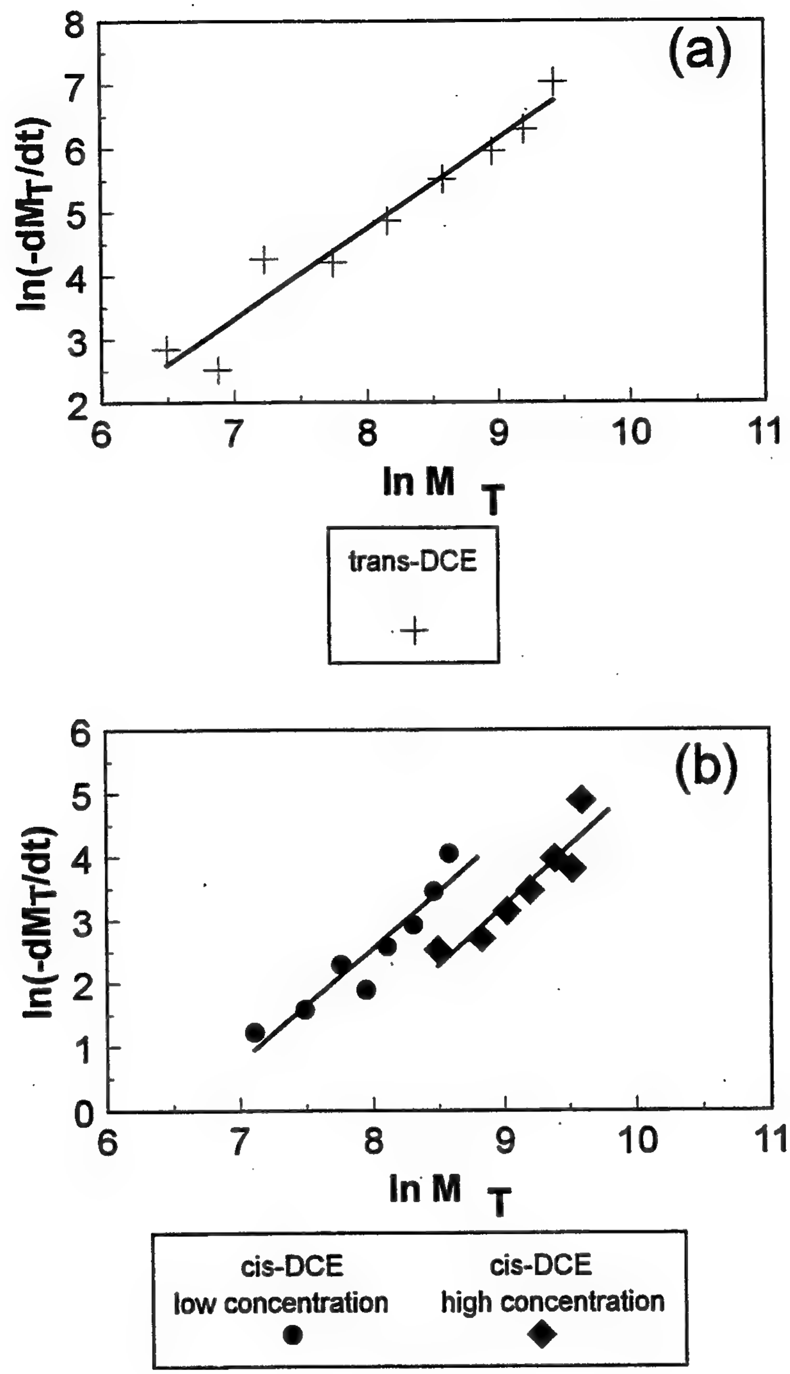


Fig. 3. Transformation rate-law determination for (a) *trans*-DCE and (b) *cis*-DCE on a total mass (M_T) basis from linearized plots, as described in the text.

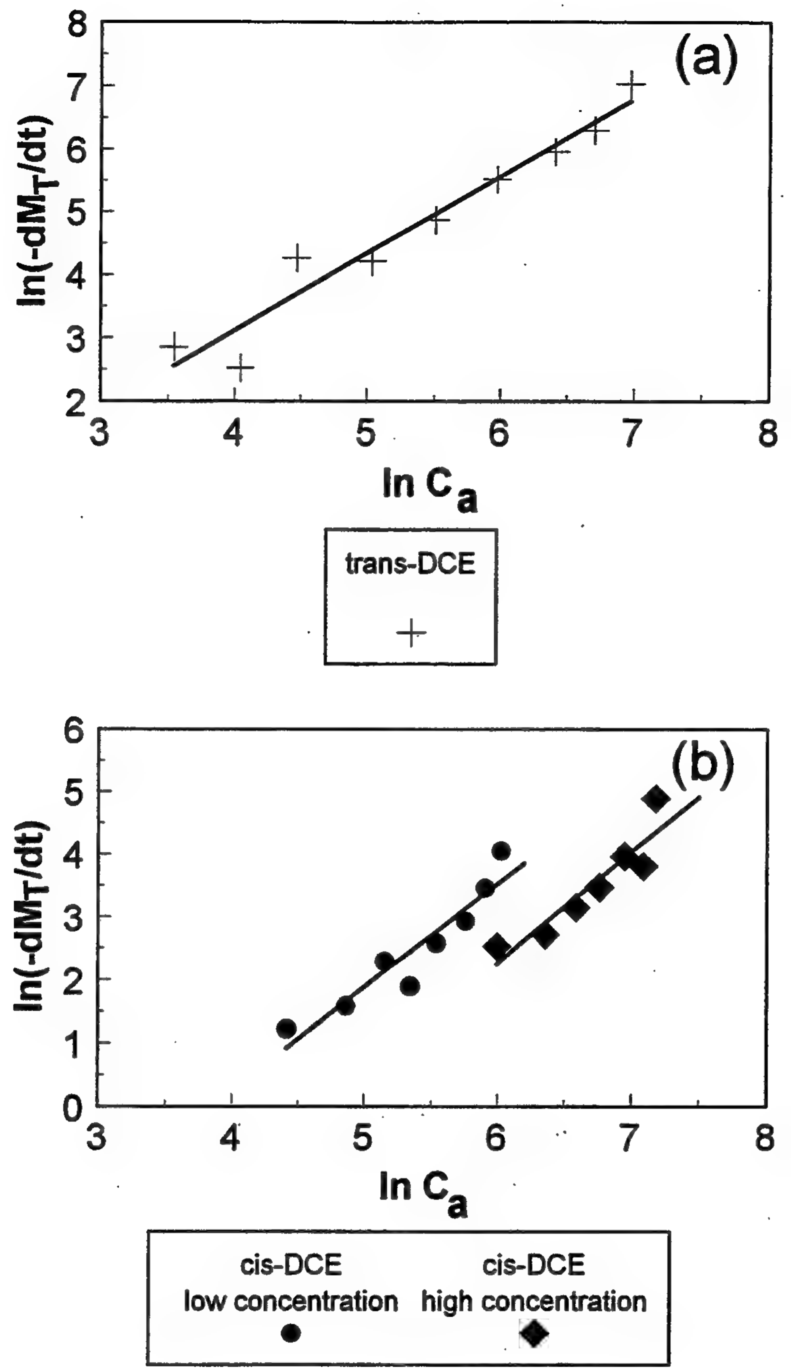


Fig. 4. Transformation rate-law determination for (a) *trans*-DCE and (b) *cis*-DCE on a solution concentration (C_a) basis from linearized plots, as described in the text.

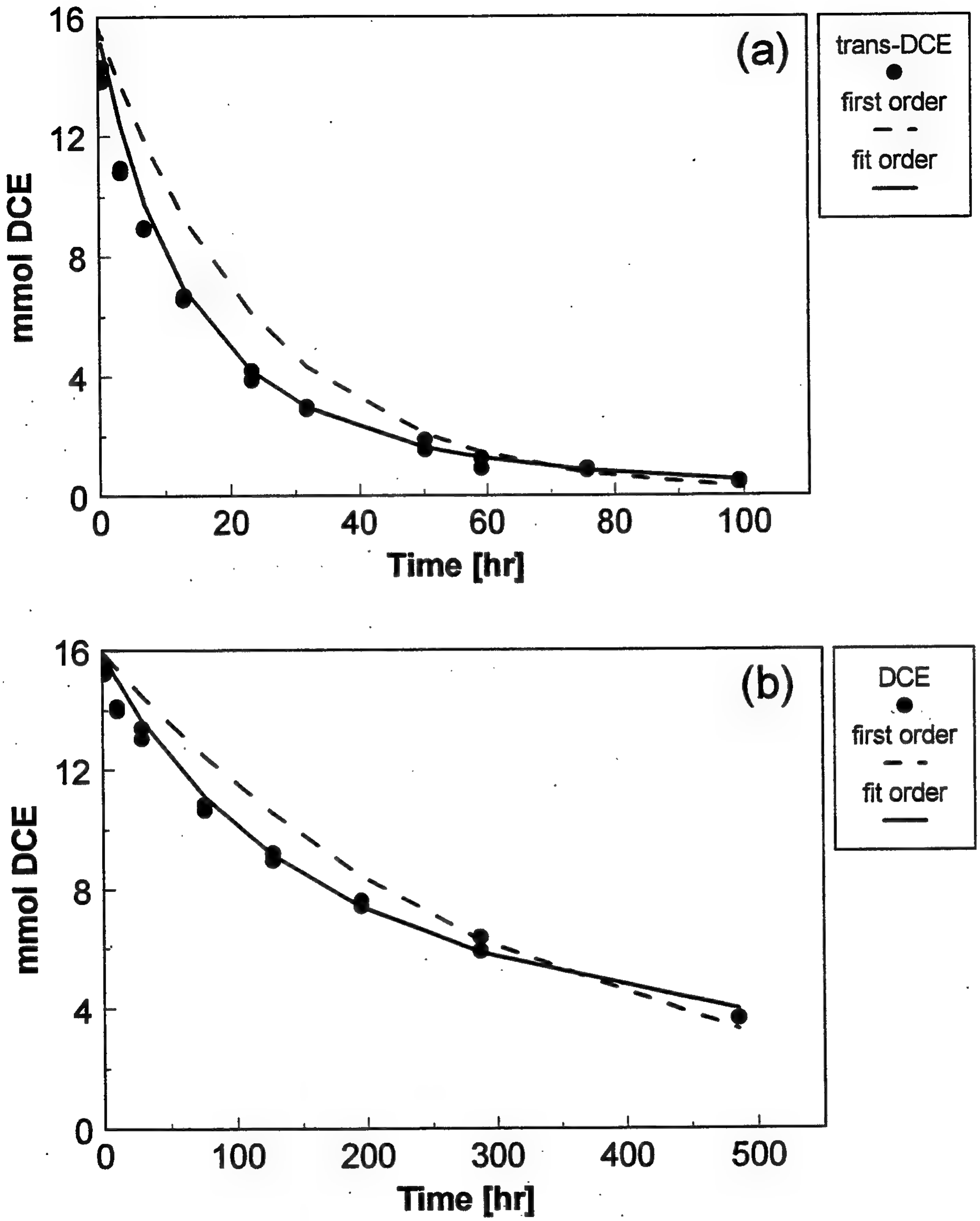


Fig. 5: Integrated first- and fit-order rates with observed total mass recovered for (a) *trans*-DCE and (b) *cis*-DCE.

Reaction products and chlorine mass balance

The organic reaction products identified near the end of the reaction course for both DCE isomers in order of decreasing predominance at the time of analysis were: ethene, ethane, C_3-C_5 alkanes, acetylene, vinyl chloride (VC). The dominant products, by far, were ethene and ethane. Vinyl chloride is transformed slowly (Campbell, pers. comm.) relative to either of the DCE isomers in this study. As a result, VC tended to accumulate in the experiment vials as the reactions progressed. Of the DCE which had been transformed at the time of measurements, approximately 2% and <0.3% had been converted to vinyl chloride for *cis*-DCE and *trans*-DCE, respectively. The fraction converted to VC was similar for both *cis*-DCE experiments.

The chlorine balance throughout the reaction course was determined by monitoring chlorine present as either DCE or free chloride ion. Free-chloride ion increased throughout the experiments (Figs. 6 and 7). The chloride background determined in DCE-free, Fe-containing controls was about 1800 nmol (140 nmol/ml or 5 mg/L). While not unimportant in the mass balance, the background was relatively low and consistent compared to the reservoirs of chlorine atoms present in the forms of DCE or free-chloride ion. The origin of the background Cl^- was either the iron metal or, perhaps more likely, the HCl used to clean the Fe. The total moles of chlorine added to each vial was estimated from the known initial DCE mass (concentration and solution volume for each vial) and the mean Cl^- background. Total chlorine observed was determined as the sum of the measured moles of DCE and Cl^- for each vial sampled. At the last sampling point shown, the measured chlorine accounts for 80 to 85% (mol/mol) of the total. The only chlorinated organic product identified was VC, which comprised a low proportion of the total chlorine added (<2%).

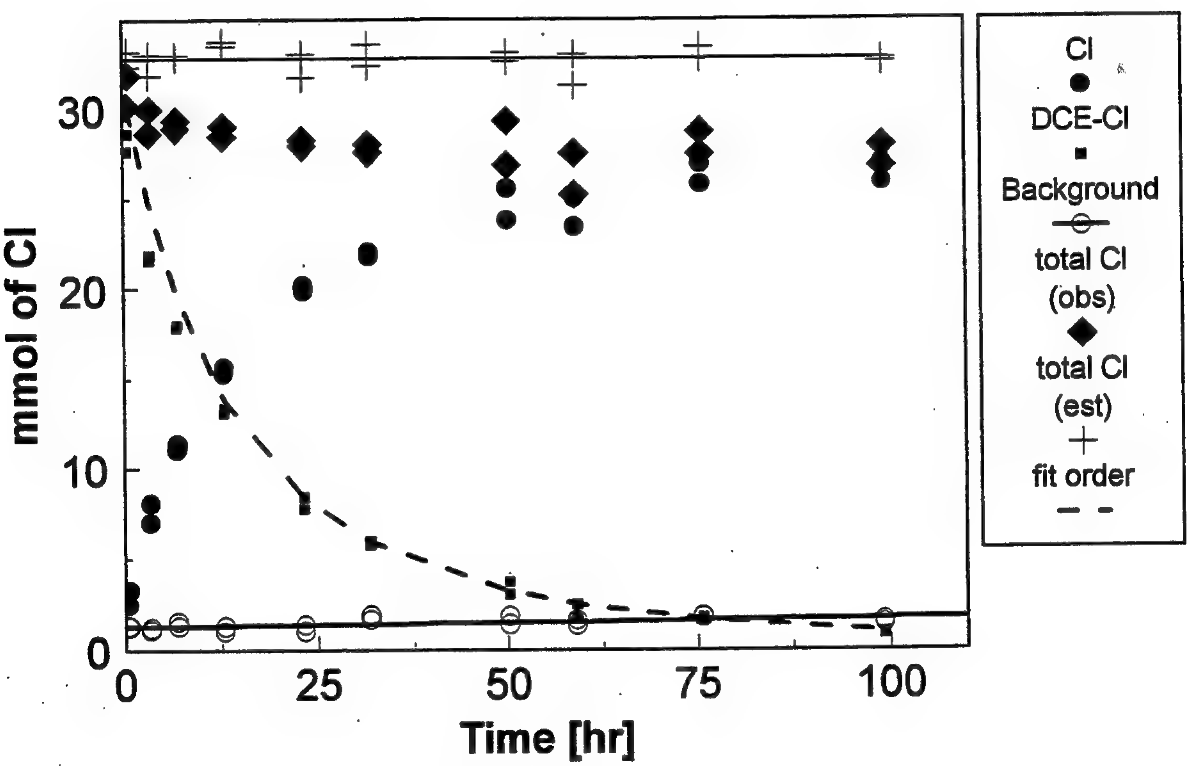


Fig. 6. Chlorine balance for *trans*-DCE. Total observed chlorine is the sum of the measured Cl^{-} and chlorine in untransformed *trans*-DCE. Total estimated chlorine is sum of the chlorine that was added initially as *trans*-DCE and the mean background Cl^{-}.

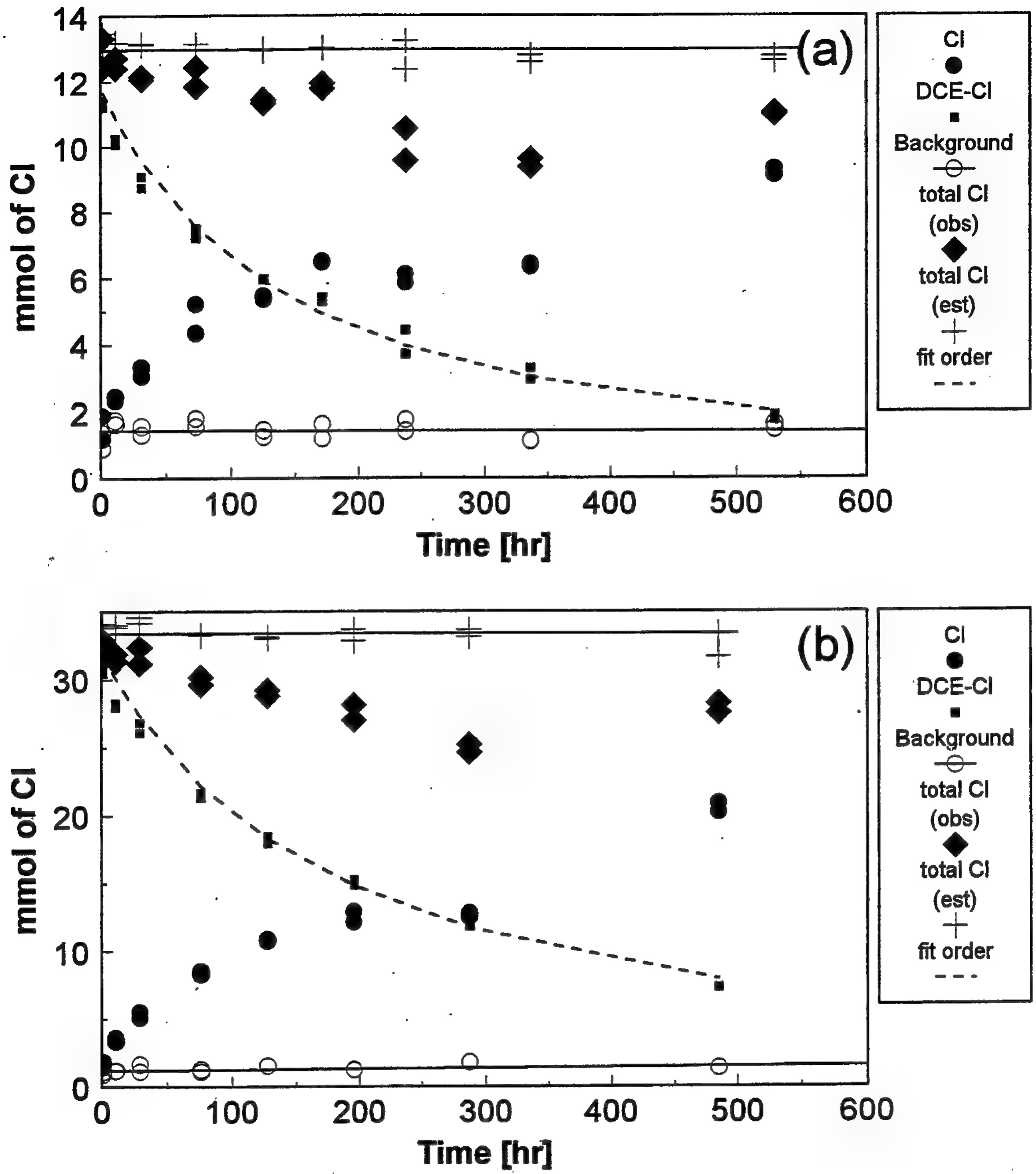


Fig. 7. Chlorine balance for (a) low concentration and (b) high concentration *cis*-DCE experiments. Total observed chlorine is the sum of the measured Cl^- and chlorine in untransformed *cis*-DCE. Total estimated chlorine is sum of the chlorine that was added initially as *cis*-DCE and the mean background Cl^-.

SUMMARY

Cis-DCE and *trans*-DCE were reductively dechlorinated in time series experiments using batch systems consisting of zero-valent iron and water. Reductive dechlorination kinetics and sorption were determined for both isomers. Unlike previous results for PCE and TCE, the transformation reaction for either isomer of DCE was not first order in solution concentration or total system concentration. Measured reaction rates and orders for the two compounds in experiments with initial concentrations of approximately 1850 nmol/ml were: 0.17 [nmol/hr]/[(nmol/ml)^{Na}] (ln λ_a = -1.79) and 0.00023 [nmol/hr]/[(nmol/ml)^{Na}] (ln λ_a = -8.37) with reaction orders 1.22 and 1.77 for *trans*-DCE and *cis*-DCE, respectively. In general *trans*-DCE was degraded much more quickly than *cis*-DCE. Chloride and organic products were produced by the transformation reaction. The chlorine mass balances for the batch systems at the end of the time series experiments were between 80 to 85%. The distribution of organic products produced by the two isomers indicates some divergence in reaction pathways. While both compounds produced large proportions of ethene and ethane, transformation of *cis*-DCE resulted in significantly greater production of vinyl chloride.

The form of sorption could be adequately described by Freundlich-type isotherms for both compounds over the concentration range measured. The Freundlich sorption coefficients (log K_F) were 0.469 and 0.471 and the Freundlich exponents ($1/n$) were 0.658 and 0.685, respectively, for *cis*-DCE and *trans*-DCE. The magnitude of sorption was greater for *trans*-DCE than the more soluble *cis*-DCE. Sorption equilibrium was apparently attained rapidly (within 1.1 hr).

REFERENCES

1. Burris, D. R., T. J. Campbell, and V. S. Manoranjan. 1995. Sorption of trichloroethylene and tetrachloroethylene in a batch reactive metallic iron-water system. *Environ. Sci. Technol.* In press.

2. Burris, D. R., V. S. Manoranjan, T. J. Campbell, and T. M. Sivavec. Tetrachloroethylene and trichloroethylene reduction by zero-valent iron: Mass transfer limitations. In prep.

3. Campbell, T. J., and D. R. Burris. 1995. Analysis of chlorinated ethene reduction products in vapor/water phase systems by dual-column, single-detector gas chromatography. *Internat. J. Environ. Anal. Chem.* In press.

4. Gillham, R. W., and S. F. O'Hannesin. 1994. Enhanced degradation of halogenated aliphatics by zero-valent iron. *Ground Water* 32:958-67.

5. Matheson, L. J., and P. G. Tratnyek. 1994. Reductive dechlorination of chlorinated methanes by iron metal. *Environ. Sci. Technol.* 28:2045-53.

6. Schwarzenbach, R. P., P. M. Gschwend, and D. M. Imboden. 1993. *Environmental Organic Chemistry*. New York, NY: John Wiley.

7. Senzaki, T. 1991. Removal of chlorinated organic compounds from wastewater by reduction process: III. Treatment of trichloroethylene with iron powder. *Kogyo Yosui* 391:29-35.

8. Sweeny, K. H. 1981. The reductive treatment of industrial wastewaters: I. Process description. In *Symposium Series 209,* ed. G. F. Bennett, 67-71. Vol 77. American Institute of Chemical Engineers.

9. Sweeny, K. H. 1981. The reductive treatment of industrial wastewaters: II Process applications. In *Symposium Series 209,* ed. G. F. Bennett, 72-78. Vol 77. American Institute of Chemical Engineers.

10. Vogel, T. M., C. S. Criddle, and P. L. McCarty. 1987. Transformations of halogenated aliphatic compounds. *Environ. Sci. Technol.* 21:722-36.

11. Westrick, J. J. 1990. National surveys of volatile organic compounds in ground and surface waters. In *Significance and Treatment of Volatile Organic Compounds in Water Supplies* N. M. Ram (ed.), 103-25. Chelsea, MI: Lewis.

12. Wilson, E. K. 1995. Zero-valent metals provide possible solution to groundwater problems. *Chemical & Engineering News,* July 3, 19-22.

RAT PUP ULTRASONIC VOCALIZATIONS: TERATOLOGIC EFFECTS OF EXPOSURE TO ULTRA-WIDE BAND ELECTROMAGNETIC RADIATION

Suzanne C. Baker
Assistant Professor
Department of Psychology

James Madison University
Harrisonburg, VA 22801

Final Report for:
Summer Faculty Research Program
Armstrong Laboratory

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, DC

and

Armstrong Laboratory

September 1995

RAT PUP ULTRASONIC VOCALIZATIONS: TERATOLOGIC EFFECTS OF EXPOSURE TO ULTRA-WIDE BAND ELECTROMAGNETIC RADIATION

Suzanne C. Baker
Assistant Professor
Department of Psychology
James Madison University

Abstract

Neonatal rats (Rattus norvegicus) typically emit ultrasonic vocalizations (UVs) when removed from their home cages and isolated from their mother and littermates. These isolation-induced UVs have been shown to be sensitive to the effects of various neuroactive substances. These vocalizations have been utilized by researchers as behavioral indicators of stress or emotionality, and they provide a useful animal model of anxiety for the investigation of the effects of various anxiogenic and anxiolytic drugs, and for teratogenic effects. This paper presents preliminary results of a study designed to examine the effects of prenatal exposure to ultra-wide band (UWB) electromagnetic radiation on rat pup isolation-induced UVs. Three groups of rat pups (prenatally exposed to lead; prenatally exposed to UWB radiation; and control) were tested for UV emission in an isolation paradigm. It was expected that the lead group would show effect of the prenatal exposure, and that (consistent with earlier studies showing no effects of UWB exposures on adults) the UWB and control groups would not exhibit significant differences in UV parameters. Findings in general were consistent with these hypotheses.

RAT PUP ULTRASONIC VOCALIZATIONS:
TERATOLOGIC EFFECTS OF EXPOSURE TO
ULTRA-WIDE BAND ELECTROMAGNETIC RADIATION

Suzanne C. Baker

INTRODUCTION

Laboratory rats (Rattus norvegicus) are commonly used as animal models in studies of the effects of drugs and other substances on both physiology and behavior. The effects of various drugs, toxic agents, and other treatments on stress and anxiety states are of particular interest. Reliable, convenient, and efficient methods of assessing psychological states in rats (such as stress, emotionality, or anxiety) are very valuable.

One behavioral pattern common to rats which has proven to be useful in this effort is vocalizations. Rats emit some vocalizations which are audible to humans. However, the majority of their vocalizations are "ultrasonic"; that is, they are above the frequency range of human sensitivity (i.e., above 20 kHz). The use of these ultrasonic vocalizations as an indicator of an animal's emotional or internal cognitive state has distinct advantages over other methods of assessing emotional state. Vocalization is a naturally-emitted behavior which occurs in response to naturally-occurring stressors in the rat's environment; the response thus has ethological validity. The vocalization response, because it is emitted naturally, does not require a training period (unlike, for example, avoidance conditioning). In addition, because the measurement of vocalization is a non-invasive procedure, it can be performed repeatedly on the same animal without interfering with the animal's ongoing behavior.

Ultrasonic vocalizations (UVs) are emitted in a variety of contexts. For example, vocalizations are emitted by adult rats in testing paradigms involving shock or acoustic startle stimulation (see Baker et al, 1991, for a

review). Rat pups typically emit vocalizations when isolated from their mother or littermates, or when roughly handled (e.g., Hofer & Shair, 1978, 1980; Carden & Hofer, 1992; Elsner, Suter, & Alder, 1990). In utero exposure to various substances (e.g., haloperidol - Cagiano et al, 1988; methylmercuric chloride - Adams, Miller, & Nelson, 1983; diazepam - Cagiano et al, 1990; nitrogen dioxide - DiGiovanni et al, 1994) has been shown to influence rat pup ultrasonic vocalizations, thus indicating the usefulness of this response in the assessment of teratogenic effects (see Elsner et al, 1988; Elsner, Suter, & Alder, 1990; Insel & Winslow, 1991). This paper reports on the results of a preliminary study examining the effects of in utero exposure of rat pups to ultra-wide band electromagnetic radiation.

RAT PUP ULTRASONIC VOCALIZATIONS

Immature rats (Rattus norvegicus) emit ultrasonic vocalizations in a variety of contexts. These vocalizations are of interest as animal models of anxiety and distress because they are influenced by anxiolytic and anxiogenic drugs as well as other substances, and because they seem to be a sensitive measure of teratologic effects (see Baker, 1994, for a review).

Neonatal rats are highly dependent on their mother. In addition to being dependent on the mother for food, the eyes and ears of newborn pups are not opened and they are unable to defecate or to thermoregulate on their own. Their locomotor abilities are very poorly developed, especially during the first week of life. For rat pups, the treatment which most reliably elicits UVs from the pup is isolation from the mother and littermates (e.g., see Carden & Hofer, 1992; Hofer & Shair, 1980; Winslow & Insel, 1991). The pup's vocalizations appear to be important in regulating the dam's behavior. In particular, they influence approach and retrieval behavior, contact, and anogenital licking of the pup by the mother (see Allin & Banks, 1972; Brouette-Lahlou et al, 1992; Smotherman et al, 1974; Smotherman et al, 1978). These vocalizations thus serve important communicative functions in the early development of the rat pup.

In typical laboratory tests of these isolation-induced UVs, pups are removed from the home cage (containing the mother and littermates) and placed in isolation for a brief period of time (typically 10 min or less). During this time, their UVs are monitored using a microphone capable of detecting sounds in the ultrasonic range.

During this "typical" isolation testing procedure, the pup is exposed to an environment which differs in many respects from its home environment. Several factors in the testing situation seem to be important in eliciting UVs. In particular, temperature, olfactory cues, and tactile stimulation have all been shown to influence rat pup UVs (see Allin & Banks, 1971; Conely & Bell, 1978; Oswalt & Meier, 1975; Lyons & Banks, 1982; Elsner et al, 1990). Of these cues, the effects of temperature have been examined in the most detail in the experimental literature. In general, the results of these studies indicate that, although other cues are important as well, placing isolated rat pups in a 20-22 deg C environment will reliably elicit UVs.

In addition to the influence of temperature, olfactory, tactile, and other cues on pup UVs, there are clear developmental changes in UVs which occur during the first 3 weeks of life. Several researchers (Noirot, 1968; Okon, 1971; Naito & Tonoue, 1987; Takahashi, 1992) report a pattern showing little to no UV emission by isolated 1-3 day old pups, peak calling rates at 5-10 days of age, then declining UV rates, particularly after 15-16 days. By about 20 days of age, UV emission in response to isolation is rare.

Some limited data are available on developmental changes in the spectral characteristics of pup calls. Naito & Tonoue (1987) reported that for pups 4-5 days old, most sound was in the range of 50 kHz. For 6-7 day old pups, energy in the 40 kHz region predominated. For pups 10 days or older, most sound was in the 30 kHz region. Takahashi (1992) reports using an ultrasonic detector ("a bat detector") tuned to 40-50 kHz for detecting calls of 7- to 14-day old pups, but tuning the detector to 30-40 kHz to detect the UVs of 21-day-old pups. These results indicate a developmental change in the predominant frequency of the calls.

PRENATAL UWB EXPOSURE - EFFECTS ON RAT PUP UV

The study reported here was designed to examine the effect of ultra-wide band (UWB) electromagnetic radiation on rat pup isolation-induced UVs. Ultra-wide band microwave exposure sources are currently being developed and tested by the USAF. Knowledge of these systems and the effects of exposure to UWB electromagnetic field is essential for the protection of AF personnel as well as for the general public. Some recently-developed exposure sources are capable of producing high-peak-power microwave (MW) pulses with relatively short pulse widths. Although it has been hypothesized that tissue damage would result from exposure to this type of radiation, the biological effects of exposure have not been thoroughly investigated (see Albanese, et al, 1994; Merritt et al, 1995). Walters et al (1995) found no effects on adult rats of acute exposure to high peak power UWB electromagnetic radiation. The study reported here will provide further information by examining the effects of longer-term, prenatal exposure on rat pups. The data reported in this paper are from a group of preliminary subjects only.

As a check on the sensitivity of the testing procedure, a group of pups who received prenatal exposure to lead was also tested using the same procedures. Although the effects of lead on rat pup UVs have not been frequently examined, it has been reported that this treatment results in an increase in UV rate (Ruoss et al, 1987). Pre- and peri-natal lead exposure has also been shown to affect human infant cries (Rothenberg et al, 1995).

Method

Subjects

Six pregnant adult females (Sprague-Dawley) were acquired in order to obtain subjects for this study. These females were randomly assigned to 3 treatment conditions (see below), so that 2 females received each treatment. After giving birth, the litters were culled so that each consisted of 4 males and 4 females. These 48 pups were the subjects for this study (16 subjects in each of the 3 treatment conditions). All pups were individually marked by

toe-clipping. Each female was housed with her offspring for the duration of the study. Animals were housed in 44 x 24 x 21 cm shoe-box-type cages, with wood chip bedding. Food (standard rodent chow) and water were available ad lib, except during testing. The animal housing room was on a 12L/12D cycle.

Apparatus

All testing equipment was located inside a sound-attenuated "testing room." Testing was done in 4 isolation chambers. (Each chamber could be fitted with a cylindrical plexiglas chamber for the measurement of acoustic startle responses (see Baker, 1993). These acoustic startle chambers were not used in the current study.) For testing, each pup was placed individually on a glass petri dish, 14 or 14.5 cm diameter, inside one of the chambers. Four pups could be tested and recorded simultaneously.

Vocalizations were detected and recorded from each chamber using an ultrasound detector ("bat detector", Ultrasound Advice Model S-25). Sensitivity of the bat detector microphone capsule is reported as better than -57 dB +/- 3 dB from 20 to 120 kHz and better than -70 dB at 180 kHz (ref 1 V/microbar). During recording, the bat detector microphone was suspended in the chamber approximately 9-10 cm above the floor of the chamber. The microphone was connected to the bat detector with a cable, and the bat detector itself was located outside the chamber, with the volume all the way down, during testing. The bat detector was used on the "frequency divide" setting, which detects signals across the entire range of sensitivity of the instrument.

Output from the bat detectors was monitored, digitized, and stored using Global Lab acquisition software (ver. 2.20) and a Data Translation A/D board (#2831-G).

Exposure Parameters

Two of the pregnant females were exposed to UWB, 2 were exposed to lead, and the remaining 2 females (the control group) received sham exposure in the UWB exposure apparatus. The lead-exposed group (LEAD) received lead in their drinking water (2000 micrograms Pb/ml lead acetate solution) from Day 2 of

gestation until they delivered their litters. Females consumed between 25 and 45 ml of solution per day, with more consumed during the later stages of gestation. After the females gave birth, regular tap water only was provided.

For the ultra-wide-band exposed group (UWB), exposures began on Day 3 of gestation. From this point, rats were exposed daily except for weekends until Day 18 of gestation, for a total of 12 days of exposure. The electromagnetic source was a Kentech (15 kV) transmitter at the Radiofrequency Radiation Division of Armstrong Laboratory. The rats were individually exposed in tapered cylindrical plexiglas holders. Exposure duration was 2 min at a pulse repetition rate of 158 MHz (major frequency) and a pulse width of 1.8 ns. Rise time was 300 ps. The peak electric field strength was 61 kV/m.

The sham-exposed females (SHAM) received the same treatment as the UWB group (including being handled and placed in the exposure apparatus) except that they received no actual UWB exposure.

<u>Procedure</u>

Day of birth of the pups was designated Day 0. Recordings were made on Days 8, 10, and 13 for all pups. This age range was chosen based on previous reports indicating that pup UVs increase over the first week of life and peak in the second week (around 10 days of age), and begin to decline rapidly after about 15 days of age. Recordings were made during the light phase of the L/D cycle. This was done in order to be consistent with other testing procedures and with tests previously done in this lab.

Several minutes prior to testing, the cage containing the female and the pups to be tested was brought into the testing room which contained the recording chambers and equipment. Pups to be tested were removed from their home cage by one of the experimenters, who picked up the pup with a gloved hand. Pups were handled as gently as possible by holding them with two or three fingers around the midsection. Sometimes it was necessary to lay the pup in the investigator's gloved hand for a brief period in order to identify the pup. The identity of the pup was noted on a data collection sheet, and the pup was placed ventral side down on a clean, room temperature glass petri

dish in one of the recording chambers. The door to the chamber was then closed. Lights in the chambers were not turned on, in order to avoid raising the temperature inside the chambers. When all 4 pups were in their chambers, recording began. Recording sessions were 6 min long. A sampling rate of 5000 Hz was used.

Following recording, pups were removed from the chambers and either placed in a holding cage for several minutes or returned to their home cage. When recording of a particular litter was completed, all the pups from that litter were placed back in the home cage with the mother and returned to the colony room. The petri dishes were washed and dried in between recording sessions. In order to avoid exposing the pups to unfamiliar odor cues, the experimenters changes gloves between litters.

Four pups were tested simultaneously until all the pups had been tested on that day. Pups from the same litter were tested at approximately the same time. Order of testing was random in other respects.

Ambient temperature in the testing room was noted throughout the recording. Temperatures ranged between 20-22 deg C.

Results

Data files generated by Global Lab were analyzed utilizing an in-house written program which rapidly detects the number of calls emitted during the session, and determines mean duration, mean intensity, and other characteristics of the signals detected. This program has been previously used in studies of adult rat acoustic-startle-induced UVs performed in this lab. The program's parameters can be adjusted to take into account the differences between adult calls and pup calls (such as shorter duration and lower intensity of pup calls), so that it maximizes detection of actual signals while minimizing "false alarms" (i.e., identifying noise in the file as a UV).

What follows is a preliminary analysis of the data. A 3 x 3 ANOVA (Treatment (LEAD vs UWB vs SHAM) x Day (DAY 8 vs DAY 10 vs DAY 13)) was

performed on the following parameters: number of calls detected; percent of time spent vocalizing; mean signal length; and mean intensity/ms. A significance level of .05 was chosen for all tests.

Data were excluded from the analyses for one subject from the LEAD group on Day 10 and for two subjects from the LEAD group on Day 13 due to equipment problems. An additional subject from the LEAD group did not vocalize on Day 13. Data for this animal were excluded from all analyses except for number of calls detected.

For the analysis of number of calls detected, there were significant main effects of both treatment ($F(2,45) = 4.27$, $p<.05$) and day ($F(2,86) = 93.69$, $p<.05$). Results of Duncan's multiple range tests indicated that animals in the LEAD group vocalized significantly more than did animals in the UWB group; however, there were no significant differences between the SHAM group and either of these other groups. Duncan's tests on the data for different days indicated that number of UVs was significantly lower on Day 13 of testing in comparison with Days 8 and 10.

For the analysis of percent of time spent vocalizing, only results for Day of testing were significant, $F(2,86) = 77.25$, $p<.05$. Results of Duncan's post-hoc tests indicated that animals vocalized most on Day 10, significantly less on Day 8, and significantly less than this on Day 13.

The analysis of mean signal length yielded a significant Treatment x Day interaction ($F(4,86)=5.34$, $p<.05$) as well as significant main effects of Treatment ($F(2,45)=3.4$, $p<.05$) and Day ($F(2,86)=7.01$, $p<.05$). Post-hoc analyses indicated that length values were significantly shorter for the LEAD group than for the UWB and SHAM groups, which did not differ from one another. Also, vocalizations were significantly longer on Day 10 than on Days 8 and 13 (which did not differ from one another).

For mean intensity/ms, there was once again a significant Day x Treatment interaction ($F(4,86)=4.21$, $p<.05$) and a significant main effect of Day ($F(2.86)=18.91$, $p<.05$) and Treatment ($F(2,45)=3.98$, $p<.05$). Vocalizations for the LEAD group had significantly lower intensity measures than the SHAM

group. The UWB group did not differ from either of these. For Day, intensity measures were highest on Day 10 of testing, significantly lower on Day 8, and lowest on Day 13.

Discussion

As was expected based on previous reports in the literature (Ruoss, et al, 1987), the lead-exposed animals emitted more UVs than did the other treatment groups. However, they were significantly different only from the UWB-exposed group and not from the sham-exposed control group. The lead-exposed animals also emitted UVs which were shorter than both the UWB and sham-exposed groups. Their UVs were also of lower intensity than the sham-exposed group (but not lower than the UWB group). There was no difference in percent of time spent vocalizing among the different treatment groups. The effect of prenatal exposure to lead on rat pup UVs thus seems to be a complex one.

Of particular interest for research in this lab is the general lack of significant differences between the UWB- and sham-exposed control groups. These two groups did not differ significantly from one another on any of the measures examined here, indicating a lack of effect of pre-natal UWB exposure on rat pup UVs, at least to the extent that can be indicated by these preliminary data.

One finding which is consistent across measures is the effect of Day of testing. The animals spent the highest percent of time vocalizing on Day 10, the vocalizations were longer, the mean intensity was greater, and the number of UVs emitted was highest. All 4 of the measures examined here generally showed an inverted U-shaped function, with vocalizing increasing from Day 8 to Day 10, and then dropping off rapidly from Day 10 to Day 13. This effect was seen regardless of treatment group. This was expected based on the published literature on the development of rat pup isolation-induced Uvs. These results also indicate the validity of the testing procedures used here in eliciting these vocalizations.

The findings from this preliminary investigation have been used to make recommendations for future testing of rat pups in this testing paradigm.

REFERENCES

Adams, J., Miller, D.R., & Nelson, C.J. Ultrasonic vocalizations as diagnostic tools in studies of developmental toxicity: An investigation of the effects of prenatal treatment with methylmercuric chloride. Neurobehavioral Toxicology and Teratology, 5, 29-34 (1983).

Albanese, R., Blaschak, J., Medina, R., & Penn, J. Ultrashort electromagnetic signals: Biophysical questions, safety issues, and medical opportunities. Aviation, Space, and Environmental Medicine, 65, A116-A120 (1994).

Allin, J.T., & Banks, E.M. Effects of temperature on ultrasound production by infant albino rats. Animal Behaviour, 20, 175-185 (1971).

Allin, J.T., & Banks, E.M. Functional aspects of ultrasound production by infant albino rats (Rattus norvegicus). Animal Behaviour, 20, 175-185 (1972).

Baker, S.C. Final Report, AFOSR Summer Faculty Research Program, Armstrong Laboratory (1993).

Baker, S.C. Rat pup ultrasonic vocalizations: A sensitive indicator of teratogenic effects. Final Report, AFOSR Summer Faculty Research Program, Armstrong Laboratory (1994).

Baker, S.C., Mulligan, B.E., & Murphy, M.R. Ultrasonic vocalizations by adult rats (Rattus norvegicus). USAF Technical Report AL-TR-1991-0135 (1991).

Brouette-Lahlou, I., Vernet-Maury, E., & Vigouroux, M. Role of pups' ultrasonic calls in a particular maternal behavior in Wistar rats: Pups' anogenital licking. Behavioural Brain Research, 50, 147-154 (1992).

Cagiano, R., Barfield, White, N.R., Pleim, E.T., Weinstein, M., & Cuomo, V. Subtle behavioural changes produced in rat pups by in utero exposure to haloperidol. European Journal of Pharmacology, 157, 45-50 (1988).

Cagiano, R., DeSalvia, M.A., Persichella, M., Renna, G., Tattoli, M., & Cuomo, V. Behavioural changes in the offspring of rats exposed to diazepam during gestation. European Journal of Pharmacology, 177, 67-74 (1990).

Carden, S.E., & Hofer, M.A. Effect of a social companion on the ultrasonic vocalizations and contact responses of 3-day-old rat pups. Behavioral Neuroscience, 106, 421-426 (1992).

Conely, L., & Bell, R.W. Neonatal ultrasounds elicited by odor cues. Developmental Psychobiology, 11, 193-197 (1978).

DiGiovanni, V., Cagiano, R., Carratu, M.R., DeSalvia, M.A., Giustino, A., & Cuomo, V. Alterations in the ontogeny of rat pup ultrasonic vocalization produced by prenatal exposure to nitrogen dioxide. Psychopharmacology, 116, 423-427 (1994).

Elsner, J., Hodel, B., Suter, K.E., Oelke, D., Ulbrich, B., Schreiner, G., Cuomo, V., Cagiano, R., Rosengren, L.E., Karlsson, J.E., & Haglid, K.G. Detection limits of different approaches in behavioral teratology, and correlation of effects with neurochemical parameters. Neurotoxicology and Teratology, 10, 155-167 (1988).

Elsner, J., Suter, K.E., & Alder, S. Microanalysis of ultrasound vocalizations of young rats: Assessment of the behavioral teratogenicity of methylmercury. Neurotoxicology and Teratology, 12, 7-14 (1990).

Hofer, M.A., & Shair, H. Ultrasonic vocalization during social interaction and isolation in 2-week-old rats. Developmental Psychobiology, 11, 495-504 (1978).

Hofer, M.A., & Shair, H. Sensory processes in the control of isolation-induced vocalization by 2 week old rats. Journal of Comparative and Physiological Psychology, 94, 271-279 (1980).

Insel, T.R., & Winslow, J.T. Rat pup ultrasonic vocalizations: An ethologically relevant behaviour responsive to anxiolytics. In: Animal models in psychopharmacology, eds. B. Olivier, J. Mos, & J.L. Slangen. Basel: Birkhauser Verlag (1991).

Lyons, D.M., & Banks, E.M. Ultrasounds in neonatal rats: Novel, predator and conspecific odor cues. Developmental Psychobiology, 15, 455-460 (1982).

Merritt, J.H., Kiel, J.L., & Hurt, W.D. Considerations for human exposure standards for fast-rise-time high-peak-power electromagnetic fields. Aviation, Space, and Environmental Medicine, 66, 586-589 (1995).

Naito, H., & Tonoue, T. Sex difference in ultrasound distress calls by rat pups. Behav. Brain Res., 25, 13-21 (1987).

Noirot, E. Ultrasounds in young rodents. II. Changes with age in albino rats. Animal Behaviour, 16, 129-134 (1968).

Okon, E.E. The temperature relations of vocalization in infant golden hamsters and Wistar rats. J. Zool., Lond., 164, 227-237 (1971).

Oswalt, G.L., & Meier, G.W. Olfactory, thermal, and tactual influences on infantile ultrasonic vocalization in rats. Developmental Psychobiology, 8, 129-135 (1975).

Rothenberg, S.J., Cansino, S., Sepkoski, C., Torres, L.M., Medina, S., Schnaas, L, Poblano, A., & Karchmer, S. Prenatal and perinatal lead exposures alter acoustic cry parameters of neonate. Neurotoxicology and Teratology, 17, 151-160 (1995).

Ruoss, E., Bruinink, A., & Elsner, J. Ultrasound vocalization, openfield, and brain-neuro-transmitter activity of rats pre- and post-natally exposed to low levels of lead carbonate. Teratology, 36, 30A-31A (1987).

Smotherman, W.P., Bell, R.W., Starzec, J., Elias, J., & Zachman, T.A. Maternal responses to infant vocalizations and olfactory cues in rats and mice. Behavioral Biology, 12, 55-66 (1974).

Smotherman, W.P., Bell, R.W., Hershberger, W.A., & Coover, G.D. Orientation to rat pup cues: Effects of maternal experiential history. Animal Behaviour, 26, 265-273 (1978).

Takahashi, L.K. Developmental expression of defensive responses during exposure to conspecific adults in preweanling rats (Rattus norvegicus). Journal of Comparative Psychology, 106, 69-77 (1992).

Walters, T.J., Mason, P.A., Sherry, C.J., Steffen, C., & Merritt, J.H. No detectable bioeffects following acute exposure to high peak power ultra-wide band electromagnetic radiation in rats. Aviation, Space, and Environmental Medicine, 66, 562-567 (1995).

Winslow, J.T., & Insel, T.R. Infant rat separation as a sensitive test for novel anxiolytics. Prog. Neuro-Psychopharmacol. & Biol. Psychiat., 15, 745-757 (1991).

DYNAMICALLY ADAPTIVE INTERFACES

Kevin B. Bennett, Ph. D.
Associate Professor
Psychology Department

Wright State University
309 Oelman Hall
Dayton, OH 45435

Final Report for:
Summer Faculty Research Program
Armstrong Laboratory

Sponsored by the
Air Force Office Of Scientific Research
Bolling Air Force Base, DC

and

Armstrong Laboratory

September 1995

ADAPTIVE INTERFACES

Kevin B. Bennett, Ph. D.
Associate Professor
Psychology Department
Wright State University

Abstract

A "dynamically adaptive interface" (DAI) is a computer interface that changes the display or control characteristics of the system (perhaps both) in real time as a function of several alternative sources of information, including system performance, contextual variables, and the current workload of the operator. The goal of dynamically adaptive interfaces is to anticipate informational needs or desires and to provide that information without the requirement for an explicit control input by the user. If properly designed DAI's have the potential to improve overall human machine system performance; they also have the very real potential to degrade performance. The fundamental challenge in designing effective DAI's is to provide dynamic changes in displays or controls that provide the right information at the right time. A collaborative research program to explore both theoretical and practical issues in dynamically adaptive interfaces has been initiated. This final report describes a conceptual framework for the development of adaptive interfaces, a DAI that has been developed to assist in precision low level navigation tasks (e.g., instrument landing), and an experimental program to evaluate that interface.

Advanced controls (a force reflecting stick) and displays (a flight director display) have been incorporated into the dynamically adaptive interface concept demonstrator. The force reflecting stick uses the haptic perceptual channel to provide feedback with regard to the optimal landing path (thus, it is not only a control, but also a display). The visual display is a redesign of the Flight Director (FD) that provides a single configural format with all information relevant to the landing task (designed to be consistent with the display principles of "correspondence" and "coherence"). To establish DAI's as a legitimate area of research and, eventually, as an effective tool for design, it must be shown that dynamically adaptive interfaces improve overall system performance. Therefore, three experimental groups are included in the proposed evaluation: the traditional interface group (conventional controls and displays always present), the advanced interface group (advanced controls and displays always present), and the dynamically adaptive advanced interface group (advanced controls and displays that are included or excluded dynamically). The timing of these adaptive changes is determined by the quality of performance, current or anticipated contextual variables, and the real-time measurement of psychophysiological indices of workload. The work completed during the SFRP represents a step towards understanding the theoretical and practical issues in a potentially useful interface design concept.

DYNAMICALLY ADAPTIVE INTERFACES

Kevin B. Bennett

Introduction

A "dynamically adaptive interface" (DAI) is an interface that changes the display or control characteristics of the system (perhaps both) in real time as a function of several alternative sources of information, including 1) performance measures, 2) the current or anticipated circumstances that exist in the system or the external world (context), and 3) the current workload of the operator. Note that this definition excludes some interfaces that one might normally consider to be adaptive. For example, many current software packages allow an individual to "personalize" their interfaces by choosing among options in a "preference" menu or window. The same capability has been implemented in other applications, including advanced aircraft. Although these interfaces are adaptive in a certain sense, the adaptation does not occur in real time. Therefore, this type of interface is excluded by the previous definition, and it is also not the focus of the current research initiative.

Changes in display or control information may also occur as a result of adaptive function allocation (e.g., Parasuraman, Hancock). Systems that have adaptive allocation dynamically transfer the responsibility for performing functions and tasks between human and machine components. Transfers in responsibility necessitate changes in the interface: at a minimum the interface must be changed to reflect which component (human or machine) is currently responsible for performing the function/task (similar to mode indication), and more substantive changes may also be required. However, once the allocation has been determined the interface remains fixed. Thus, although adaptive function allocation is closely related to adaptive interfaces, it is also not the focus of the current research initiative.

The increasing sophistication and capability of a broad spectrum of computer technologies have made the concept of dynamically adaptive interfaces a very real possibility. DAI's have the potential to improve overall performance of the human-machine system dramatically through an increased capability (relative to static interfaces) to provide the right information, in the right format, at the right time. The dominant theoretical perspective on human computer interface design describes effective interfaces as those which have achieved "transparency" (Hutchins, Hollan, and Norman, 1986). That is, the interface effectively disappears, thus enabling the user to interact directly with the objects of interest in the domain, and to achieve effective interaction a minimum of cognitive effort. Adaptive interfaces have the potential to take transparency one step farther. In traditional interfaces appropriate control inputs must be provided by the user when additional information is needed or desired; in dynamically adaptive interfaces this need or desire will be anticipated and the relevant information will be provided without the requirement for control input by the user.

Despite the potential for dynamically adaptive interfaces to facilitate human machine performance there has been very little research and development in the area. This is probably due to the fact that the defining characteristic of DAI's (adaptivity) appears to violate one of the fundamental principles of effective interface design. Both practical

guidelines and theoretical approaches to interface design identify consistency as a critical component. When individuals are presented with consistent information they are able to develop extremely effective patterns of behavior over time, behavior that is characterized by the parallel and automatic processing of both external information and internal responses. With fairly trivial nuances to differentiate these terms, researchers have labelled this type of behavior as "automatic processing" (e.g., Shiffrin and Schneider, 1977), "skill-based behavior" (Rasmussen, 1986), "procedural knowledge" (Anderson, 1982), and "associative skill" (Fitts and Posner, 1967).

Thus the fundamental challenge in designing effective DAI's is to provide dynamic changes in display or control information that do not interfere with either the development or the execution of this type of behavior. If designed properly, DAI's have the potential to improve overall human machine system performance by anticipating informational needs and providing that information in a timely fashion. If designed improperly, DAI's have the potential to degrade system performance by preventing automatic behaviors from developing, by presenting irrelevant information, or in the worst case by eliminating information that is needed. The key to developing effective dynamically adaptive interfaces is to ensure that the dynamic changes in the interface are consistent with both the current goals of the user and the current context of the system.

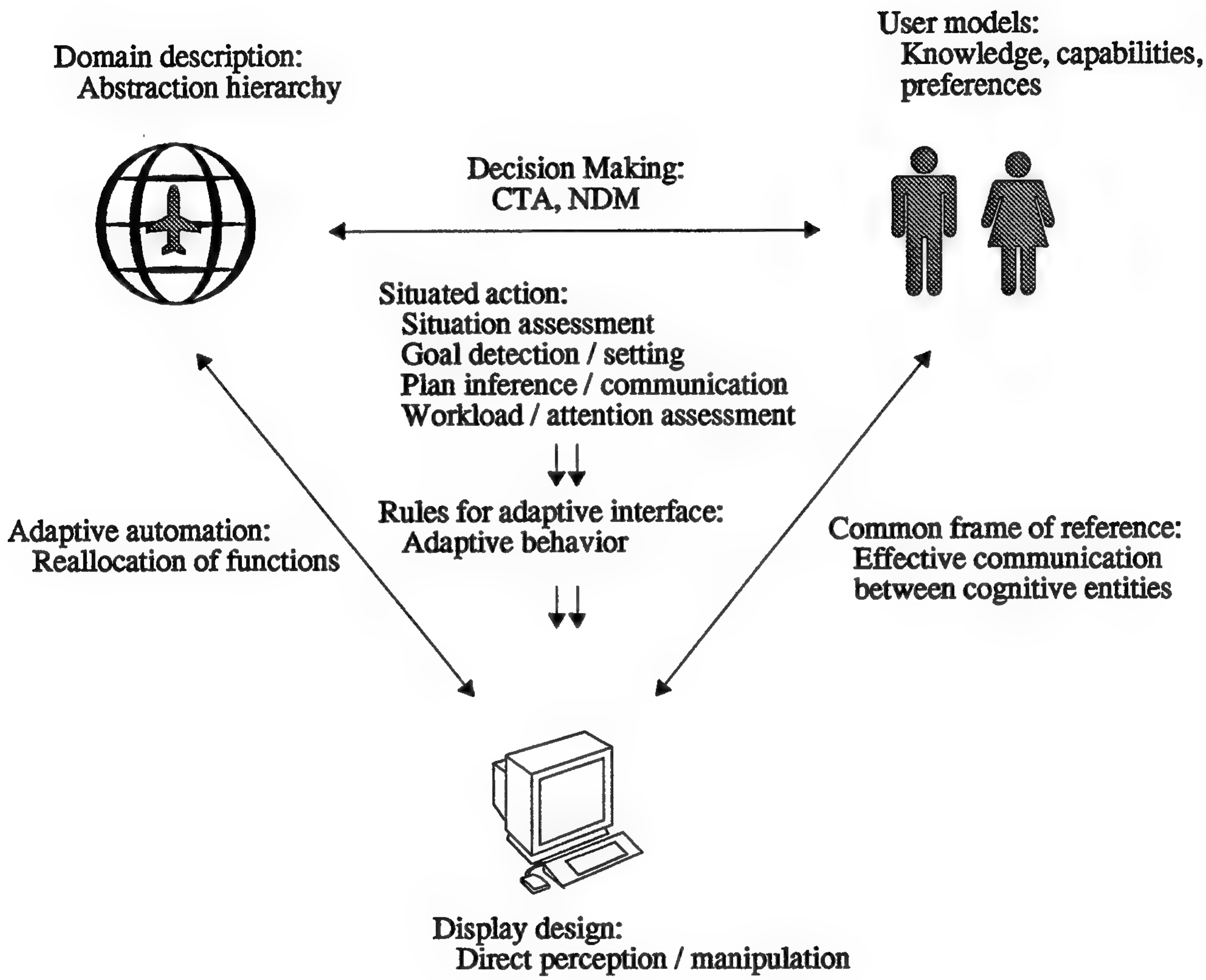
The Summer Faculty Research Program has provided the opportunity to begin a collaborative research initiative to explore both theoretical and practical issues in dynamically adaptive interfaces. Progress in both areas is a necessary prerequisite to the development and implementation of DAI's. An appropriate theoretical perspective is required to guide decisions about the choice of dynamic behaviors, about the information and knowledge that should be used to trigger these adaptive behaviors, and about the orchestration of these behaviors and information sources to facilitate performance. The theoretical aspects will benefit tremendously from the experience gained in actually implementing a DAI. This final report describes a conceptual framework for the development of adaptive interfaces, a DAI that has been developed to assist in precision low level navigation tasks (e.g., instrument landing), and an experimental program to evaluate that interface.

Theoretical perspective of dynamically adaptive interfaces

A theoretical approach referred to as "cognitive systems engineering" (CSE, Rasmussen, 1984) has been applied to the practical problem of developing DAI's. A component of this approach, the cognitive system triad (Woods and Roth, 1988), will be used to frame the discussion. The cognitive triad assumes that the quality of performance in complex, dynamic domains is the result of three interactive and mutually constraining components (see Figure 1): the cognitive demands produced by the domain of interest, the cognitive agent(s) that meet those demands, and the representation of the domain through which the agent experiences and interacts with the domain (the interface). From this perspective an adaptive interface must be able to recognize that a cognitive demand / cognitive resource mismatch has occurred (i.e., that the cognitive demands produced by the domain have exceeded the cognitive resources that the agent has available to meet them). After recognizing a demand/resource mismatch an adaptive interface must be able to determine the appropriate change in the amount or type of information that is required to alleviate the mismatch, and

Figure 1.

Conceptual Framework for Adaptive Interfaces



to adapt accordingly. There are many sources of knowledge that are potentially relevant to these adaptive behaviors and several of these sources will be discussed in greater detail.

Knowledge of user

User models. One potential source of information that might be used to initiate changes in the interface is related to the knowledge, capabilities or preferences of a user. User-specific models could be devised to represent 1) an individual's knowledge of the system (e.g., capabilities, limitations, tendencies that are specific to the interface and its adaptive behaviors), 2) an individual's knowledge of the task domain (e.g., declarative or procedural knowledge), or 3) other information about a particular user (e.g., processing capabilities, interaction style, preferences, strategies). In theory, these models could be used as the basis for changes in the interface. However, the implementation and use of these models is still in an exploratory phase, most notably in the area of intelligent tutoring systems (e.g., Polson and Richardson, 1988). Given the complexity, cost of implementation, and questionable utility of these models this is not a reasonable option for the proposed research initiative.

Workload. A second source of information about the user that might be useful in the implementation of DAI's is related to the level of workload that an individual is experiencing. It is a fairly well established fact that the relationship between workload and performance is not necessarily a linear one. More specifically, equal increases or decreases in workload are not reflected by equal increments or decrements in performance. As workload increases from a low to a high level individuals are often able to mobilize resources to meet the increased demand, and performance may not suffer. However, as an individual approaches the maximum level of workload small increases may have precipitous (and negative) effects on performance (e.g., the "straw that broke the camel's back"). From the perspective of dynamically adaptive interfaces, on-line psychophysiological measurements of workload might be useful through the provision of information that could be used to improve the timing of adaptive interface changes, and therefore overall performance. In particular, on-line workload measurements could provide a source of information that complements other measurements. As the previous analysis suggests, performance may be maintained acceptable levels with high workload, and rules for adaptation that are based on these performance levels might not be invoked. However, levels of performance that approach boundary conditions when combined with high levels of workload could be a useful indication that interface changes should be implemented.

Focus of attention. One additional source of information about the user that could be used to initiate changes in the interface is information about an individual's focus of attention. On-line assessments of line of sight (including eye and head tracking) could provide information regarding the activities and information that an individual is currently seeking.

Knowledge of domain

Developing effective interfaces requires a deep understanding and explicit description of the "semantics" of a work domain. This requirement is even more important for developing dynamically adaptive interfaces. Rasmussen's

abstraction hierarchy (1986) is a theoretical framework for describing domain semantics in terms of a nested hierarchy of functional constraints (including goals, physical laws, regulations, organizational/structural constraints, equipment constraints, and temporal/spatial constraints). One way to think about the abstraction hierarchy is that it provides structured categories of information (i.e., the alternative conceptual perspectives) that an individual must consider in the course of accomplishing system goals. Thus, in complex domains, situation awareness requires the operator to understand the process at different levels of abstraction. Further, the operator must be able to understand constraints at one level of abstraction in terms of constraints at other levels.

Knowledge of decision making in domain.

An abstraction hierarchy analysis provides a description of the domain constraints, independent of decision making in the domain. The design of effective adaptive interfaces will also require a complementary analysis: a cognitive task analysis. The overall goal of this analysis is to describe decision making within a domain. Cognitive task analysis provides an understanding and description of decision making within constraints imposed by domain, including procedures, strategies, steps, subgoals, required information, and interrelations between procedures. Historically, decision research has focused on developing models that describe the generation of multiple alternatives (potentially all alternatives), the evaluation (ranking) of these alternatives, and the selection of the most appropriate alternative. By and large, perception was ignored. In contrast, recent developments in decision research, stimulated by research on naturalistic decision making (e.g., Klein, Orasanu, and Zsambok, 1993) has begun to give more consideration to the generation of alternatives in the context of dynamic demands for action. Experts are viewed as generating and evaluating a few "good" alternatives. The emphasis is on recognition (e.g., how is this problem similar, or dissimilar, to problems that I have encountered before?). As a result, perception plays a dominant role. This change in emphasis has increased awareness of perceptual processes and dynamic action constraints in decision making. These trends have, either directly or indirectly, led researchers in interface design to focus on the representation problem. Thus, the challenge for display design from this perspective is to provide appropriate representations that support humans in their problem solving endeavors.

Interface Design

Recently, a number of research groups have recognized that effective interfaces depend on both the mapping from human to display (the coherence problem) and the mapping from display to a work domain or problem space (the correspondence problem). Terms used to articulate this recognition include direct perception (Moray, Lee, Vicente, Jones, and Rasmussen 1994) ecological interface design (Rasmussen and Vicente, 1989) representational design (Woods, 1991), or semantic mapping (Bennett and Flach, 1992). Collectively, these principles complement the principles of direct manipulation articulated by Hutchins, Hollan, and Norman (1986). There are two critical components of effective display design: correspondence and coherence.

Correspondence. Correspondence refers to the issue of content --- what information should be present in the interface in order to meet the cognitive demands of the work domain? Correspondence is defined neither by the domain itself,

nor the interface itself: it is a property that arises from the interaction of the two. Thus, in Figure 1 correspondence is represented by the labelled arrow that connects the domain and the interface. One convenient way to conceptualize correspondence is as the quality of the mapping between the interface and the work space, where these mappings can vary in terms of the degree of specificity (consistency, invariance, or correspondence).

Coherence. Coherence refers to the mapping from the representation to the human perceiver. Here the focus is on the visual properties of the representation. What distinctions within the representation are discriminable to the human operator? How do the graphical elements fit together or coalesce within the representation? Is each element distinct or separable? Are the elements absorbed within an integral whole, thus losing their individual distinctness? Or do the elements combine to produce configural or global properties? Are some elements or properties of the representation more or less salient than other elements or properties? In general, coherence addresses the question of how the various elements within a representation compete for attentional and cognitive resources. Just as work domains can be characterized in terms of a nested hierarchy of constraints, so to, can complex visual representations be perceived as a hierarchy of nested structures, with local elements combining to produce more global patterns or symmetries.

The mapping problem. Whether a display will be effective or not is be determined by both correspondence and coherence. More specifically, the effectiveness of a display is determined by the quality of the mapping between the constraints that exist in the domain and the constraints that exist in the display. The display constraints are defined by the spatio-temporal structure (the visual appearance of the display over time) that results from the particular representation chosen. In configural representations the geometrical display constraints will generally take the form of symmetries --- equality (e.g., length, angle, area), parallel lines, colinearity, or reflection. The core problem in implementing effective displays is to provide visual representations that are perceived as accurate reflections of the abstract domain constraints: Are the critical domain constraints appropriately reflected in the geometrical constraints in the display? Are breaks in the domain constraints (e.g., abnormal or emergency conditions) reflected by breaks in the geometrical constraints (e.g., emergent features such as non-equality, non-parallelism, non-closure, bad form)? Only when this occurs will the cognitive agent be able to obtain meaning about the underlying domain in an effective fashion.

Dynamically adaptive interfaces

A central component of dynamically adaptive displays is that they anticipate informational needs and adapt without explicit control inputs from the operator. This is a form of automation (the system is in control of the initiation and nature of the informational changes) and can be contrasted to other forms of interaction including user controlled (the operator requests the informational change) or variably controlled (either the human or the computer can be in control). Under these circumstances the interface must be considered as a "cognitive agent" which is capable of both intelligent and unintelligent action. For dynamically adaptive interfaces intelligent action consists of the adaptive changing of information that is consistent with knowledge about the user, their workload, the task at hand, and the context as defined by the current state of the domain. Unintelligent action would consist of the failure to provide information that

is relevant, or to take away information that is currently needed.

The problem of coordinating intelligent machine and human activities has been investigated recently, primarily in the context of systems that have a machine expert system (e.g., Roth, Bennett, and Woods, 1987; Suchman, 1987). Two central principles have emerged: "situated action" and "mutual intelligibility" (which depends upon a "common frame of reference"). Suchman (1987) has proposed that human-human communication provides a particularly relevant analogy to frame questions of human computer interaction. She contrasts the traditional view of intelligent action (the development and implementation of plans) to situated action, stating that "... purposeful actions are inevitably *situated actions*. By situated actions I mean simply actions taken in the context of particular, concrete circumstances" (p. viii). Suchman applies this view of intelligent action to the design of human computer systems, and observes that "Interaction between people and machines implies mutual intelligibility, or shared understanding" (p. 6). Roth, Bennett, and Woods (1987) reached similar conclusions in their evaluation of an expert system designed to assist technicians in the repair of an electro-mechanical device. The design of the system interface was "opaque" and therefore inhibited the development of a mutual understanding between the human and machine experts. As a result the two cognitive entities worked independently and in parallel (rather than orchestrating their activities), and overall system performance was degraded significantly.

The preceding theoretical perspective, and the analyses that it suggests, forms the basis for the development of dynamically adaptive interfaces. The abstraction hierarchy analysis reveals the critical domain constraints (the cognitive demands that must be met, and the domain resources that are available to meet them). The cognitive task analysis defines the decisions that need to be made to meet domain goals, and the information that is relevant to those decisions. This information is used, in conjunction with the semantic mapping principles of display design, to develop displays that appropriately reflect domain constraints and thereby assist in decision making and problem solving.

These analyses also provide knowledge about situated action, and constitute the basis for dynamically adaptive behaviors on the part of the interface. Applying these analyses will result in a definition of what adaptations are appropriate in the interface, and the development of rules that describe when those adaptations should occur. These rules include knowledge of the operator's performance, the operator's workload, general aspects of the task/domain, and specific aspects of current system state (i.e., an assessment and continuous monitoring of the domain for changes that have implications for goals and required actions -- the context or situation). The end result will be a common frame of reference, or mutual intelligibility, between the human and the adaptive interface. Dynamic alterations in the interface will be consistent with current goals and context; the intelligent action on the part of the interface will improve overall human-machine system performance.

Dynamically adaptive interface for low level navigation tasks

One very important category of aviation tasks is referred to as low level navigation. A characteristic of low level navigation tasks is the requirement to fly an aircraft along a predetermined path (or, at a minimum, to intersect

predetermined way points). This requirement may be accompanied by the need to be at a specific point in the path at a specific time. One example of a low level navigation task is to deliver ordnance in enemy territory. To accomplish this a pilot may be required to fly along a particular path (to avoid ground based threats), and may also be required to arrive at way points or the target site at a particular point in time (to benefit from air cover that has been provided to mask arrival, or to coordinate with other offensive activities).

A dynamically adaptive interface was developed in the context of a particularly difficult low level navigation task: instrument landing. Both advanced controls (a force reflecting stick) and displays (a flight director display) have been incorporated into the dynamically adaptive display. The force reflecting stick provides changes in resistance to a pilot's control input (or the amount of force that is required to implement the control input) that varies as a function of the airplane's deviation from the optimal approach path. The flight director display integrates several pieces of landing-related information (e.g., glide slope, line-up information, airspeed) in a single "configural" display that provides a commanded steering input to the pilot. The presence or absence of both the stick and the configural display constitute the dynamic adaptive behavior of the interface. The dynamic allocation of this information will be determined by on-line measurements of performance at the landing task and by on-line measurements of workload (psychophysiological). The dynamically adaptive interface (controls, displays, adaptive behaviors, rules for adaptation) and the task environment (landing segment, performance measurements) will be described in this section.

Background: Force-reflecting feedback for instrument landing system

The proposed research initiative will complement ongoing research and development efforts at the Human Interface Technologies Branch of Armstrong Laboratories. Brickman, Hettinger, Roe, Lu, Repperger, and Haas (1995) developed and investigated a force-reflecting, haptically-augmented aircraft control stick designed to improve performance at a simulated instrument landing task. When the ground surface is obstructed (e.g., when flying through low clouds), a pilot relies upon instruments that provide information with respect to an optimal approach path to the runway (in particular, glideslope and line-up information). The force reflection stick represents an augmentation of existing instrumentation. In contrast to a conventional stick, the augmented stick serves as both a control and a display. For example, if the plane deviated to the right of the optimal approach path the pilot would experience an increase in resistance when attempting control inputs to the right, and a decrease in resistance when attempting control inputs to the left. Thus, the force reflecting stick is not only a control, but also a display which uses the haptic channel to provide feedback with regard to the optimal approach path.

Brickman et. al.'s evaluation of this augmented stick were generally favorable. Objective measurements of performance at the landing task and subjective estimates of workload were obtained under a factorial combination of stick (force-reflecting or regular), turbulence (present or absent), and initial displacement from runway (five positions). Performance measurements included variables related to the quality of the landing (e.g., lateral and longitudinal RMS error) and control inputs to the stick (e.g., velocity, power, and energy). In general, the force-reflecting stick improved objective measurements of performance at the task, and decreased subjective measurements

of workload.

However, for the purposes of the proposed dynamically adaptive interface the most interesting finding of the study was found in some non-significant trends. For two measurements (lateral RMS error and subjective work load ratings) it was found that the force-reflecting stick improved performance and decreased workload under high turbulence conditions, but degraded performance and increased workload under low turbulence conditions. These differences did not reach statistical significance, and the pattern was not present for all measurements. However, the results suggest that pilots may benefit from the dynamic inclusion and exclusion of force-reflective feedback information in the interface: overall performance at the landing task may be improved if force-reflective haptic feedback is provided when the landing task is difficult (e.g., high turbulence) and removed when it is not difficult.

Force reflection stick

The force reflection stick used in the Brickman et. al study will be modified to provide tactile feedback in both dimensions.

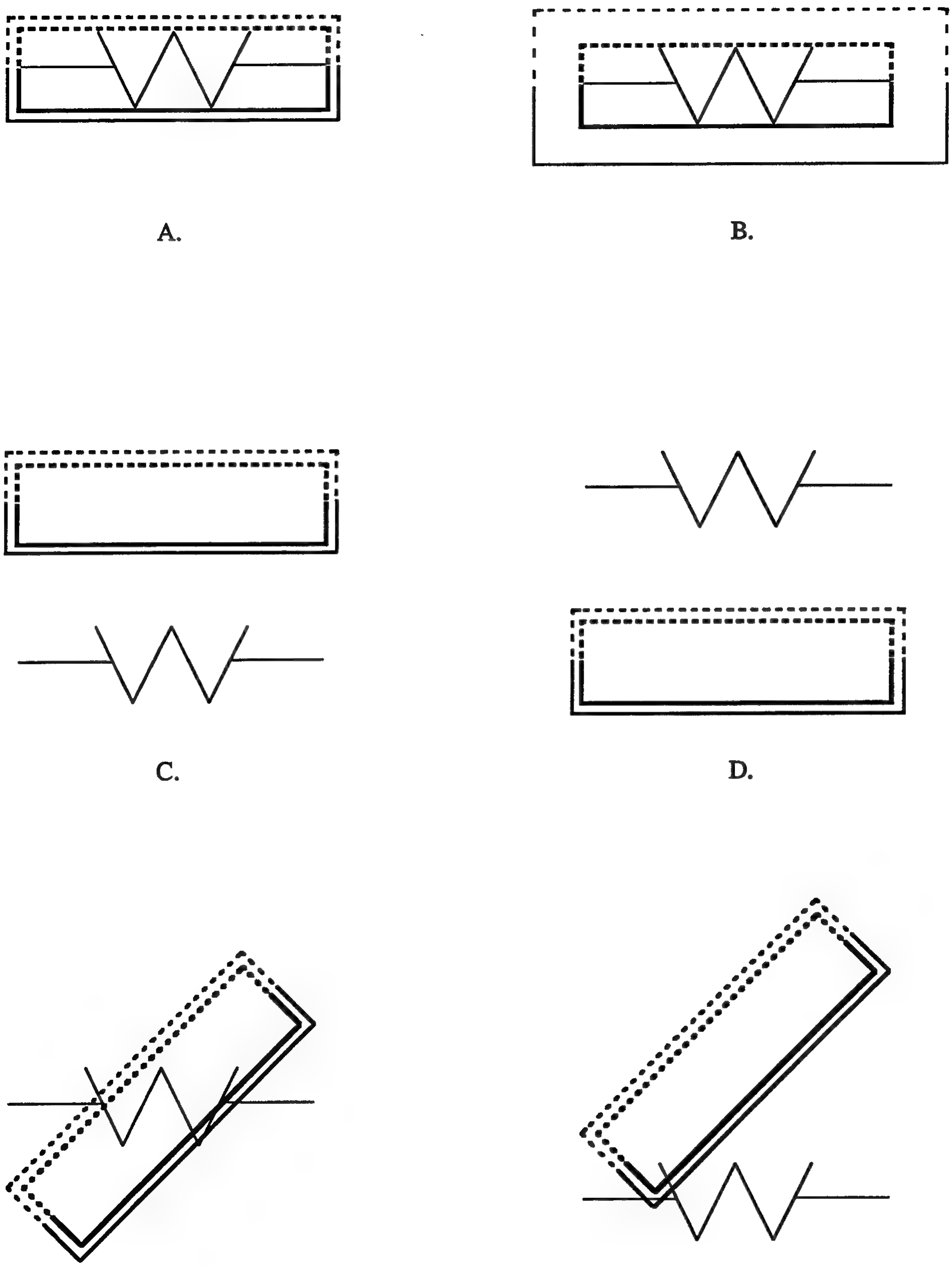
Visual Display: Flight director

The visual display component of the DAI is a redesign of the Flight Director (FD). The modifications were based on the theoretical principles outlined in Bennett and Flach (1992) and Bennett, Toms, and Woods (1993). Thus, the goal of the display is to provide a single configural display that presents all information relevant to the landing task (the principle of "correspondence") in a centralized and easily interpretable format (the principle of "coherence"). The display integrates information about heading, roll, pitch, altitude, and airspeed in a single integrated configural display (see Figure 2). It does not present absolute values for these variables. Instead, the information presented represents a recommended control input(s) to the pilot, relative to the optimal approach path. The display suggests the appropriate control input through a coherent representation that integrates all of the relevant variables that should be considered in the decision.

The display consists of two rectangular boxes (see Figure 2a) and a visual reference point (the watermark symbol). The visual reference point remains in a fixed position in the HUD and both rectangles move dynamically to signify deviations from the recommended landing path. Both rectangles have a dashed and a solid component. The solid component serves as a reference to ground, while the dashed component serves as a reference to the sky. This aspect of the display serves as a cue for the aircraft-ground relationship: when the plane is right side up the solid portion will appear on the bottom, when the plane is upside down the solid portion will appear on the top.

The bold and the non-bold rectangles are used to represent deviations in actual and recommended airspeed. The size of the bold rectangle is a visual representation of the goal value for airspeed during landing, which is 150 knots. Since the recommended airspeed does not change during an approach, the size of this rectangle will remain constant. The non-bold rectangle provides a visual representation of current airspeed, and the size of this rectangle will vary accordingly. The relationship between the two rectangles provides an "emergent visual feature" that specifies progress

Figure 2



towards obtaining the airspeed goal. When current airspeed is greater than 150 knots the non-bold rectangle will be larger than the bold rectangle (see Figure 2b); when current airspeed is less than 150 knots the non-bold rectangle will be smaller than the bold rectangle. Thus, both the goal for airspeed and deviations from this goal are represented directly in the display.

Changes in the location and orientation of these two rectangles (relative to the fixed visual reference point) provide additional emergent features that signify deviations in altitude, heading, pitch, and roll directly. If all variables are consistent with the optimal landing path then the rectangles will enclose the watermark symbol and will be centered and aligned with it (see Figure 2a). A deviation in altitude (pitch) is represented by a vertical displacement of the rectangles. When the rectangles are above the watermark the airplane is below the recommended flight path (see Figure 2c); when the rectangles are below the watermark the plane is above the recommended flight path (see Figure 2d). A deviation in heading is represented by rotation of the rectangles. When the airplane's course is to the right of the recommended flight path the rectangles will rotated counter-clockwise, relative to the watermark (see Figure 2e), and vice-versa. A deviation in heading and pitch is represented by a rotation and a vertical displacement (see Figure 2f). This figure represents a commanded input to turn the plane to the left and to simultaneously increase altitude.

The general visual characteristics of the FPM display, as described in the previous paragraphs, will change under certain conditions. The first and second conditions involve the airplane's relative position in the optimal approach path. As the airplane approaches the runway (on the final approach segment) the visual representation of a deviation in the display will change. To facilitate landing performance the display will become progressively more "sensitive" as the airplane approaches the runway in the final segment. For example, a deviation in altitude at the beginning segment of the optimal approach path will produce less vertical displacement than the same deviation in the final segment.

A second change in the visual characteristics of the display is related to the pilot's decision to continue or abort the landing. As the airplane approaches the runway the pilot must evaluate all relevant variables to determine if the associated values are within acceptable parameters for continuing the landing. The decision point represents the last possible instance that this decision can be made, because there will not be sufficient time to implement control inputs. When this decision point is reached the rectangles will begin to blink rapidly, emphasizing that if the variables are not within acceptable parameters, then the landing should be aborted immediately.

The third condition is related to extreme glide slope, roll, or heading deviations. Under extreme conditions the FPM will approach the boundaries of the HUD. Although a portion of the FPM will be allowed to disappear, it will never disappear completely. A fraction (perhaps 1/5) of the appropriate portion of the display will remain at the boundary of the HUD.

Dynamically adaptive behavior

Performance boundaries. One source of information that will be used as a basis for the dynamic adaptation of the

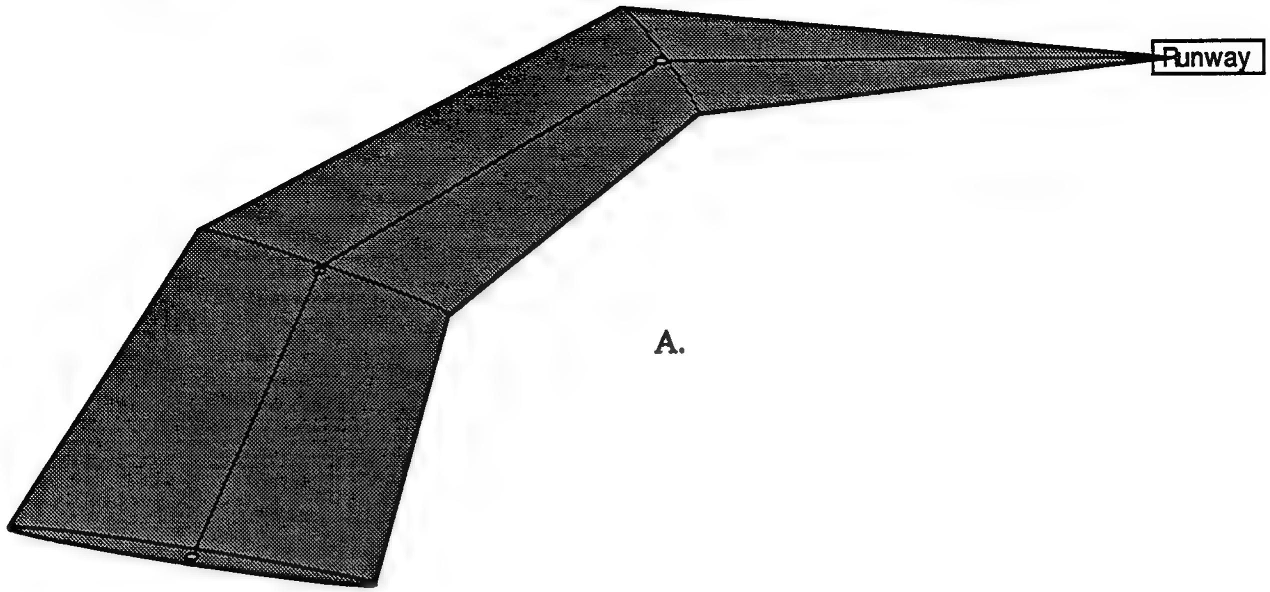
interface is performance based. The airplane's position relative to the optimal path will be monitored, and the degree of deviation from this path will serve as a primary indication that information needs to be added to, or taken from, the interface. One way to conceptualize the optimal path is as a line that exists in three-dimensional space. Similarly, the performance boundaries can be conceptualized as geometrical shapes that enclose this line in three-dimensional space. In cross section the performance boundaries will be shaped as an ellipsis that is elongated in the horizontal axis. This geometrical shape is required because in the landing task deviation in altitude is deemed to be more important than deviation in bearing. Thus, the general shape of the performance boundaries is an elliptically-shaped "cylinder" that surrounds the optimal path in three-dimensional space (see Figure 3). The shape of these geometric performance boundaries will become more restrictive for each segment of the path. Figure 3a represents a top-down perspective on the performance boundaries; Figure 3b represents a "face-on" view; Figure 3c represents a side view.

There will be several performance boundaries that will act in concert as one determinant of adaptive behavior on the part of the interface. There will be at least two nested cylinders representing performance boundaries (see Figure 3). Maintaining the position of the aircraft within the inner cylinder will result in no changes to the standard interface. When aircraft position breaches the inner performance boundary the force-reflective properties of the control stick will become apparent in the interface. When aircraft position breaches the second performance boundary the FPM display information will be presented in the interface (in addition to force-reflection). Similarly, if performance improves (if aircraft position moves within either of the performance boundaries) the associated information will be removed from the interface.

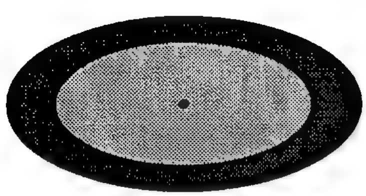
This final aspect of dynamic adaptive interface will bear scrutiny. It is here that the potential for violation of consistency, and the negative implications for performance, are most likely to appear. It may be that the pilot is able to improve landing path position within a performance boundary, but with great difficulty. Therefore, it may not be advisable to remove the additional interface information immediately, in case the performance is hovering at the edge of the performance boundary. One possibility is to implement a "dead band" for the removal of either FPM or force-reflection information that allows the information to remain until the pilot achieves a higher degree of acceptable performance. The concept of the dead band would also eliminate a related effect -- informational "hysteresis" when the performance variable oscillates on either side of the performance boundary and information is added, and taken away in quick succession. Perhaps this dead band would be a function of both the amount of improved performance (the distance from the performance boundary) and the amount of time that performance has been at that level.

<u>Psychophysiological measures of workload</u>. The real-time measurement of psychophysiological indices of workload will be used as a complementary knowledge source to assist in defining the timing of dynamic adaptations to the interface. The performance measures and associated rules outlined previously will be the primary basis for adaptation when an individual is experiencing low levels of workload. However, when an individual is experiencing high levels of workload these criteria will be modified. The goal is to anticipate performance decrements that could result from high workload situations, and to modify the timing of dynamic changes in the interface to decrease the probability of

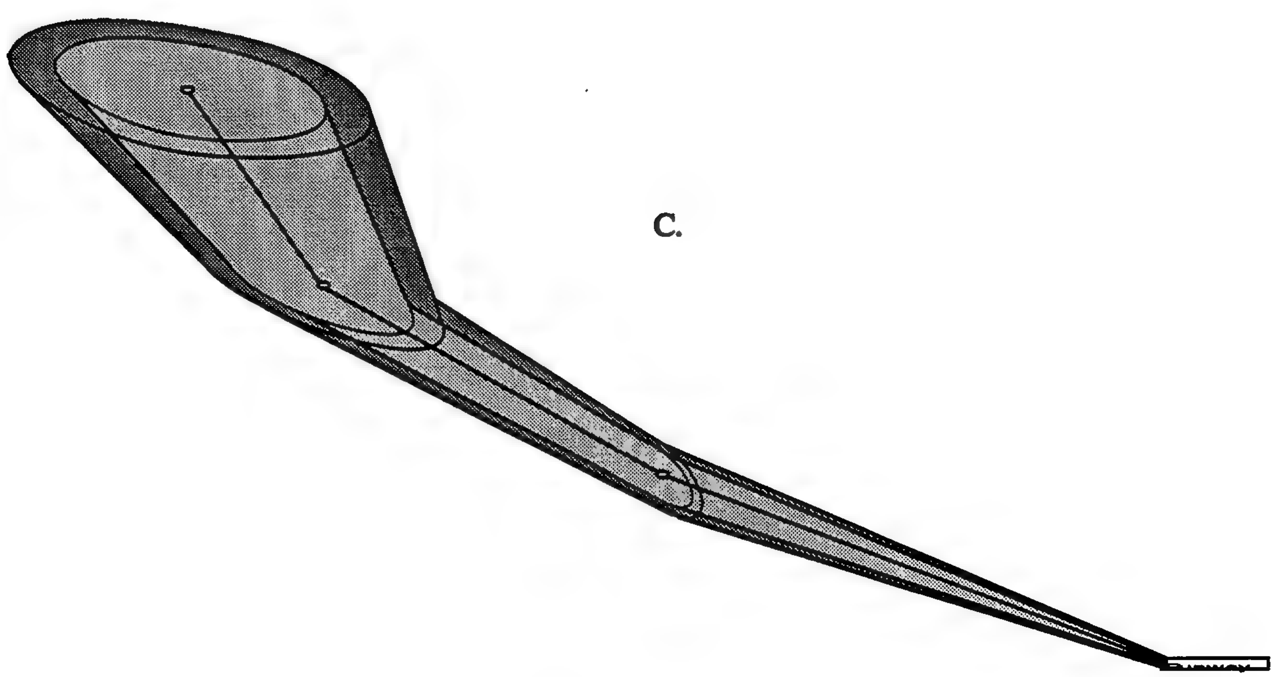
Figure 3



A.



B.



C.

these performance decrements. For example, when an individual is experiencing high levels of workload a smaller deviation from the optimal approach path will trigger the addition of a display. Similarly, a smaller deviation from the optimal approach path will be required for the removal of a display if the individual is still experiencing high levels of workload. One way to conceptualize this is as two different sets of performance boundaries that apply under high and low workload conditions (additional performance boundaries are not shown in Figure 3).

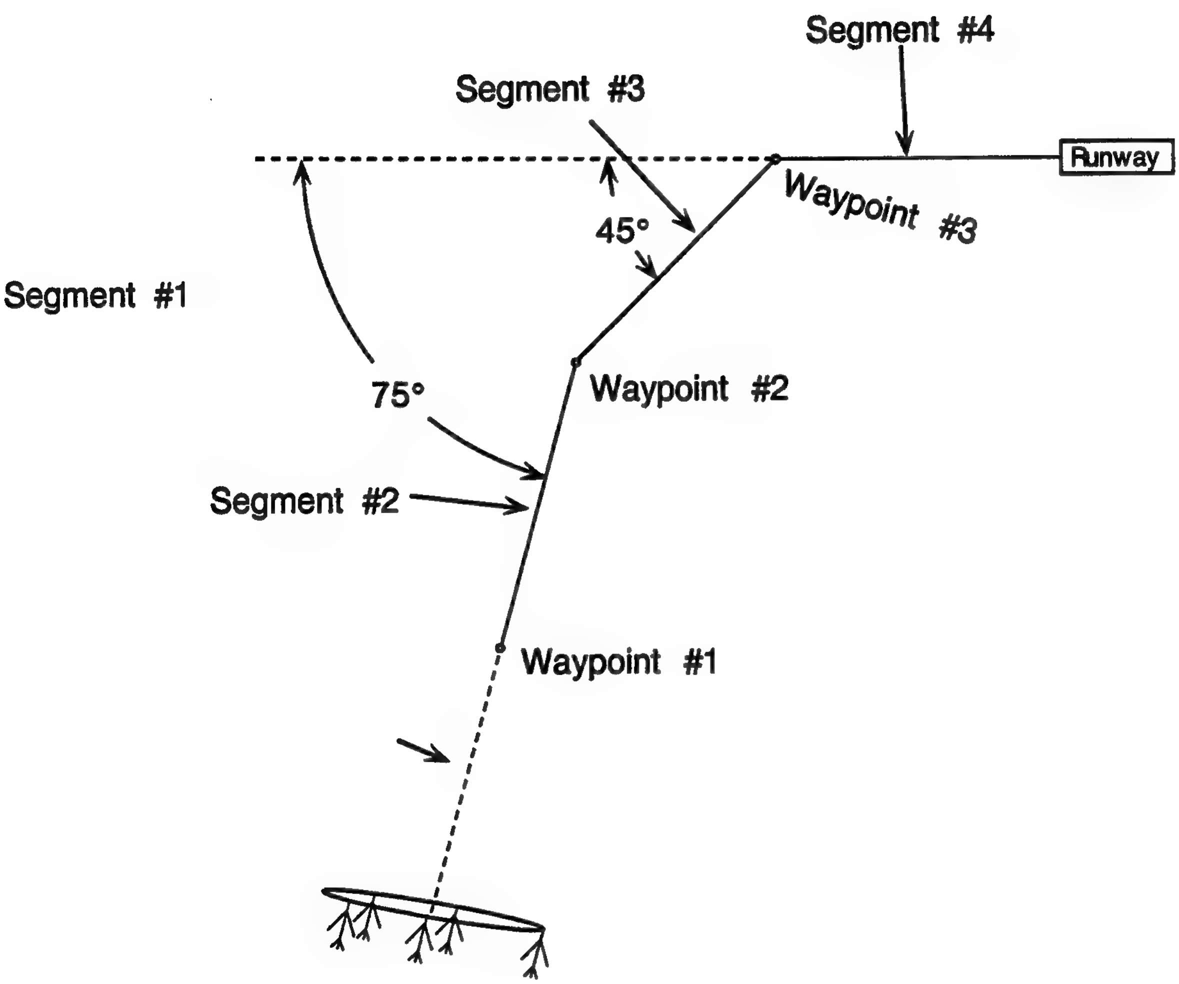
The Workload Assessment Monitor (WAM) will provide on-line measurements of workload. WAM was developed at the Flight Psychophysiology Laboratory of Armstrong Laboratory. It has two major functions: to collect and reduce psychophysiological data and to classify the operator's workload state in real time. Heart beat, eye blink, respiration, and brain wave (seven channel EEG) activity is detected and measured (including inter-event intervals, amplitudes, and durations). WAM uses these data to compute estimates of subjective workload continuously, using a Bayesian quadratic statistical classifier. Before its use in the proposed experiment WAM will be calibrated with respect to the instrument landing task. Brickman et. al. assessed workload using the NASA TLX, which was completed by each subject after each experimental trial. These measures revealed significantly different subjective impressions of workload, primarily related to the degree of turbulence that was encountered. WAM will be calibrated by replicating the experimental conditions and comparing the subjective measurements of workload provided by WAM and the NASA TLX estimates.

Proposed evaluation

To establish DAI's as a legitimate area of research and, eventually, as an effective tool for design, it must be shown that dynamically adaptive interfaces can improve overall system performance. This possibility will be investigated using an expanded version of the instrument landing task used by Brickman et. al. In the ILS simulation the optimal approach path consisted of a single heading and a two mile approach segment. To allow the opportunity for adaptive interface behavior the optimal approach path will be longer (16 miles) and will have multiple segments (with different headings and associated way points, see Figure 4). Five initial starting points will be used in the study, each the same distance (four miles) from the first waypoint. These starting points will be distributed so that they are equidistant from the first waypoint. The heading (015), pitch (0), and roll (0) of the plane will be the same in all five starting positions. These five starting positions will require initial control inputs on the part of the pilot to intercept the first waypoint. The first four mile segment of the MLS course will have a heading of 015 degrees. At the second waypoint the MLS course changes to a heading of 065 degrees (a fifty degree change in heading). The optimal path to negotiate all changes in heading will be defined as an arc connecting the two course segments, with the waypoint in the middle of the arc. The arc will represent an optimal turn using a 30 degree bank angle at 150 knots. The third segment will have a heading of 065 degrees, and therefore represents a 45 degree intercept heading to the final approach segment, which has a heading of 110 degrees. The altitude of the optimal approach path at any point will be determined by applying a 3 per cent glide slope.

Approximately thirty pilots will serve as experimental subjects, and will be randomly assigned to three experimental

Figure 4



Segment #4
Segment #3
Runway
Waypoint #3
45°
Segment #1
75°
Waypoint #2
Segment #2
Waypoint #1

groups: traditional interface, advanced interface, and dynamically adaptive advanced interface. The traditional interface group will perform the ILS task with an interface that approximates the information that is currently available, including glide slope and localizer deviation displays located in the head-down ADI display. The advanced interface group will perform the task with the heads-up flight director display and the force-reflective control stick that were described earlier. These advanced controls and displays will always be present in the interface. For the final experimental group the interface will adapt dynamically: using the logic outlined previously the advanced controls and displays will appear or disappear as a function of performance, work load, or contextual information.

In addition to the between group variable of interface, three within group manipulations will be used. The first of these variables is the amount of turbulence that is present during an experimental trial. The high and low turbulence conditions employed by Brickman et. al will also be used in the proposed study. The second within groups variable will be the initial starting position for a trial (five initial starting positions for an individual trial, see Figure 4). The final within groups variable will be experimental session (10 sessions). Thus, in each of 10 experimental sessions a subject will complete 20 landing trials, a factorial combination of turbulence (2 levels) and starting position (5 levels). The order of trial presentation will be randomly determined.

Several independent variables will be gathered. In the real world it is often imperative to stay as close to the prescribed landing path as possible for a variety of reasons (e.g., coordination with other aircraft, threat avoidance). To assess the overall quality of performance on this dimension an RMS error score will be calculated for each trial. This score will assess the relationship between the optimal landing path and the actual landing path. A second set of dependent variables will assess another critical consideration in the real world, the quality of touchdown performance. These dependent variables will include crashes, and both lateral and longitudinal deviation from the optimal touchdown location on the runway. A final set of dependent variables will assess the nature of the subjects' control stick inputs, including stick velocity, power imparted to the stick, energy imparted to the stick, and power integrated over time. A 3 (interface group) x 2 (turbulence) x 5 (starting position) x 10 (experimental session) mixed ANOVA will be performed for each of these dependent variables.

<u>Rationale and predicted experimental outcomes</u>

The advanced interface was designed to provide improved control and display information, relative to the traditional interface. Therefore, it is predicted that this interface will facilitate performance at the ILS task. Because the pilots in the experiment will be much more familiar with the traditional interface, and because ten hours represents a relatively short learning period, this is a strong prediction. If this pattern of results is obtained it will be a strong validation of the utility of the displays and controls that were developed, as well as the underlying theoretical perspective that was used to develop them. In terms of the overall goal of the proposed research program -- to investigate theoretical and practical issues in the development of dynamically adaptive interfaces -- the more interesting experimental outcome will be the contrast between relative levels of performance that are produced by the advanced and the dynamically adaptive advanced interfaces. To be a useful concept in interface design it must be demonstrated that an adaptive

version of an interface can produce increments in performance, relative to a static version that portrays the same relative information. The proposed experiment represents a direct test of this requirement. In Brickman et. al's results there were non-significant trends suggesting that the force-reflective stick facilitated performance under high turbulence conditions, but hindered performance under low turbulence conditions. Based on these results a reasonable initial prediction is that performance between the two interfaces will be reasonably equivalent under high turbulence conditions, but that the adaptive interface will facilitate performance under low turbulence conditions. Based on the results of this initial experiment, and available time and resources, additional experiments may be conducted.

Acknowledgments

I would like to thank those individuals and organizations who have contributed their time, effort, and resources to making the SFRP appointment successful and mutually beneficial. First, I would like to thank the Air Force Office of Scientific Research, Research & Development Laboratories, the Human Engineering Division of Armstrong Laboratories, and the Human Interface Technologies Branch for making the appointment possible. I would also like to thank the contractors who assisted in both the conceptual and physical development of the concept demonstrator (Bob Shaw, Larry Hettinger, Bart Brickman, Liem Lu, Barbara Palmer, David Snider). Finally, I would like to extend a special thanks to Michael Haas, the Director of the Human Interface Technologies Branch, for his enthusiastic support in all phases of the appointment and project development.

References

Anderson, J. R. (1982). Acquisition of cognitive skill. Psychological Review, 89, 369-406.

Bennett, K. B., and Flach, J. M. (1992). Graphical displays: Implications for divided attention, focused attention, and problem solving. Human Factors, 34(5), 513-533.

Bennett, K. B., Toms, M. L., and Woods, D. D. (1993). Emergent features and configural elements: Designing more effective configural displays. Human Factors, 35(1), 71-97.

Brickman, B. J., Hettinger, L. J., Roe, M. M., Lu, L., Repperger, D. W., and Haas, M. W. (In Press). Haptic specification of environmental events: Implications for the design of adaptive, virtual interfaces. To appear in Proceedings of VRAIS 1996. New York, NY: IEEE

Corbett, J. M. (1985). Prospective work design of a human-centred CNC lathe. Behaviour and Information Technology, 4 (3), 201-214.

Fitts, P. M., and Posner, M. I. (1967). Human performance. Belmont, CA: Brooks Cole.

Hutchins, E. L., Hollan, J. D., and Norman, D. A. (1986). Direct manipulation interfaces. In D. A. Norman, and S. W. Draper (Eds.), User centered system design (pp. 87-124). Hillsdale, NJ: Lawrence Earlbaum Associates.

Klein, G. A., Orasanu, J., and Zsambok, C. E. (Eds.). (1993). Decision making in action: Models and methods. Norwood, NJ: Ablex Publishing Corp.

Moray, N., Lee, J., Vicente, K.J., Jones, B.G., and Rasmussen, J. (1994). A direct perception interface for nuclear power plants. Proceedings of the Human Factors and Ergonomics Society 38th Annual Meeting (pp. 481-485). Santa Monica, CA: Human Factors and Ergonomics Society.

Polson, M. C., and Richardson, J. J. (1988). Foundations of intelligent tutoring systems. Hillsdale, NJ: Lawrence Earlbaum Associates.

Rasmussen, J. (1986). Information processing and human-machine interaction: An approach to cognitive engineering. New York, NY: Elsevier Publishing Co., Inc.

Rasmussen, J., and Vicente, K. (1989). Coping with human errors through system design: Implications for ecological interface design. International Journal of Man-Machine Studies, 31, 517-534.

Roth, E. M., Bennett, K. B., and Woods, D. D. (1987). Human interaction with an 'intelligent' machine. International Journal of Man-Machine Studies, 27, 479-526.

Shiffrin, R. M. and Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. Psychological Review, 84, 127-190.

Suchman, L. A. (1987). Plans and situated actions: The problem of human machine communication. New York, NY: Cambridge University Press.

Woods, D. D. (1991a). The cognitive engineering of problem representations. In G. R. S. Weir and J. L. Alty (Eds.), Human-computer interaction and complex systems (pp. 169-188). London: Academic Press.

Woods, D. D., and Roth, E. M. (1988). Cognitive systems engineering. In M. Helander (Ed.), Handbook of human-computer interaction (pp. 1-41). Amsterdam: Elsevier Science Publishers B. V. (North-Holland).

A STUDY OF X-RAY FLUORESCENCE ANALYSIS OF LEAD IN PAINT

Verne L. Biddle
Chemistry Professor
Department of Chemistry

Bob Jones University
1700 Wade Hampton Boulevard
Greenville, SC 29614

Final Report for:
Summer Faculty Research Program
Armstrong Laboratory
Brooks Air Force Base

Sponsored by:
Air Force Office of Scientific Research
Bolling AFB, Washington DC

and

Armstrong Laboratory

July 1995

A STUDY OF X-RAY FLUORESCENCE ANALYSIS OF LEAD IN PAINT

Verne L. Biddle
Chemistry Professor
Department of Chemistry
Bob Jones University

Abstract

The assay of lead in paint samples was studied, using a Kevex XRF 770 Spectrometer and a Kevex Analyst 8000 data processor. Relative standard deviations and accuracy were significantly better for ground paint than for paint chips. The mass of the sample in the XRF cup was determined to not be extremely important, as long as it was not much less than about one gram. A least squares calibration curve, obtained from a series of ELPAT paint samples, was used for quantitation.

Analysis of standard ELPAT paint samples using XRF gave very good results, as did the analysis of spiked samples. The detection limit was well within the range necessary to ascertain lead content as it pertains to lead abatement programs. A number of field samples of paint were analyzed and compared with XRF analysis of the paint chips and with the standard ASTM digestion-AA method. Multilayered paints exhibited the greatest differences for the various methods, especially if they contained significant amounts of lead. Based on the good correlation with known paints analyzed simultaneously as method checks, XRF analysis of ground paint (particles passing a 100-mesh sieve) was shown to be as good as or superior to the presently used AA method. Also demonstrated was the importance of grinding the paint sample, even for the AA method Software that was PC-based was found to be more easily used, faster, and more adaptable to the present application than was the software presently in use in the Analyst 8000.

A STUDY OF X-RAY FLUORESCENCE ANALYSIS
OF LEAD IN PAINT

Verne L. Biddle

Introduction

The presence of lead-based paint in hospitals, schools, other public buildings, and play equipment, for example, has received relatively recent scrutiny due to its health hazard. Since lead is both toxic and teratogenic, it is important that the lead content of paint be accurately and efficiently measured at levels in the region of the maximum acceptable limit of 0.06 wt%. This is particularly relevant to lead abatement programs. The currently accepted ASTM method of analysis for lead in dried paint samples (ES 37-94) involves acid digestion of the sample, aided by microwave energy, followed by atomic absorption assay of the solution produced.

The aforementioned ASTM method tends to be time-consuming and therefore does not permit rapid throughput of samples. Furthermore, it has been noted in this laboratory that the acid digestion does not always seem to be complete, which could lead to erroneous results. The purpose of the present study was to determine the feasibility and reliability of using X-ray fluorescence (XRF) as a quantitative method for lead in paint, and, if it was, to devise a procedure for such an analysis.

Since the flux of the X-ray tube used in XRF analysis changes as it ages and even from day to day, any quantitation method must have a method of correcting for these intensity changes. Furthermore, complex matrices, such as are often found in environmental samples, may absorb the incident or the emitted (fluoresced) X rays to varying degrees, or they may exhibit secondary fluorescence. For these reasons, a method of quantitation whereby the ratio of the analyte peak intensity to that of an internal standard or monitor peak is usually used. For this study, the reference peak was the incoherent scatter peak (Compton peak) from the silver secondary target used for exciting the sample.

Preliminary Studies

Several variables concerning the paint samples had to be investigated. However, due to the fact that the XRF spectrometer that was originally to be in place for this study--the Kevex 770/Analyst 8000 system--was not assembled until well into the research period, many of these preliminary investigations regarding sample handling were performed on the Kevex 700 spectrometer using PC-based software by IXRF Systems, Inc., for data processing.

The Analyst 8000 uses the XRF Toolbox™ software for the analyzer to provide the control functions for data acquisition, spectral display, element identification, and data processing. This software is a compiled FORTRAN program that runs on IOMEGA Bernoulli disk drives containing two 10 Mb-cartridges. The manual supplied by Kevex contains an extensive section on the meaning of the many commands available to the user for programming and instructing the software to carry out the desired functions. An extensive amount of time was needed in this study to learn the software sufficiently to

operate the system and to write a program that would automate the analysis of a tray of samples without operator intervention. This program and instructions for its use appear in the Appendix.

The Kevex 700, on the other hand, is controlled by a *much* more easily learned software package that is based on Windows™, uses a mouse and icons, and can be learned rather quickly. Further discussion of the two systems will appear later in this report.

Reproducibility. The first variable to be investigated was the reproducibility of results using the samples in the form of chips, approximately 0.5- to 1-cm in diameter. Between each run, the samples were removed from the cup and replaced. The results are summarized in Table 1. It is apparent that the samples that had the highest percent relative standard deviation (RSD) were those that were multilayered and had more than minor amounts of lead. This observation could be explained by the differing composition of the layers composing them. That is, results would vary with the number of chips having the higher lead-containing layer(s) facing the X-ray beam.

Table 1. Reproducibility for Analysis of Paint Chips

Sample Number	*Average Wt% ± RSD (%)*[a]	*Description*	*Multilayered?*
34442	0.0099 ± 0.0024 (24%)	thin coating on sheetrock facing	no
35048	3.84 ± 1.19 (31%)	—	yes
35049	0.0096 ± 0.0009 (9%)	very thin coating on paper	no
35050	0.032 ± 0.003 (9%)	slightly flexible, thin flakes/sheets	no
35051	9.82 ± 6.11 (62%)	slightly brittle, plaster (?) backing	yes
35052	0.196 ± 0.048 (24%)	brittle	yes
35053	0.065 ± 0.001 (15%)	brittle, some small pieces	yes
35054	3.85 ± 1.12 (29%)	flexible, multicolored	yes(?)
35055	6.37 ± 2.21 (35%)	thick, brittle	yes
35056	0.0087 ± 0.0013 (15%)	irregular thickness, on caulk(?)	yes
35190	0.098 ± 0.025 (26%)	multicolor, some small pieces	yes

[a]Results of 7 analyses

This was further investigated by comparing results for several paints when the chips were randomly oriented with the results when the layers were manually oriented with either the face or the backside down. The results from this preliminary observation clearly indicated that the orientation of multilayered paint chips containing lead had a sizable effect on the results. For a more definitive investigation, two paints were chosen which had the following characteristics: 1) available in sufficient quantities, 2) multilayered, 3) brittle for ease of grinding. One appeared to consist of at least three layers (35187) and one consisted of six identifiable layers (35188). The results appear in Table 2. In the first two conditions, randomly oriented paint chips were on top of several layers of manually oriented paint chips. Between each analysis in the randomly oriented condition, the samples were dumped out of the cup and

returned. The results seemed to clearly indicate that some form of sample grinding was necessary in order to obtain reproducible and accurate results.

Table 2. Importance of Orientation of Multilayer Paint Chips

Sample Number	*Backside Down [wt% Pb ± RSD (%)]*[a]	*Frontside Down [wt% Pb ± RSD (%)]*[a]	*Randomly Oriented [wt% Pb ± RSD (%)]*[a]
35187	0.095 ± 0.012 (13%)	0.049 ± 0.026 (53%)	0.159 ± 0.024 (15%)
35188	2.06 ± 0.05 (2.4%)	0.379 ± 0.025 (6.6%)	1.63 ± 0.34 (21%)

[a]Results of 3 analyses

Particle Size. Since grinding of the sample appears to be necessary, the particle size to be used must be determined. Paint samples were ground with a mortar and pestle since no mill was available at the time of this study. Sample cups for analysis were filled about one-fourth full of ground paint. Six runs were made for each of two samples at each mesh size, giving a total of twelve analyses for each mesh fraction. Between each run, the cup contents were dumped out, mixed, and returned. The results can be seen in Table 4, which reveals that the relative percent standard deviation decreases with decreasing particle size (i.e., with increasing mesh number; see Table 3). Further, the signal increased with decreasing particle size, as would be predicted from better packing of particles and thus more paint being excited by the X rays. The poorer reproducibility of the larger particle fractions is expected since the proportion of particles with a given layer ("face") facing the X ray will vary randomly. Smaller particles do not have "faces" at the macroscopic level.

Table 3. Relationship of Mesh Number and Particle Size

Mesh Size	*Particle Size (µm)*
10	2000
20	841
40	425
100	149

Table 4. Reproducibility and Particle Size

	Average Observed Wt% Pb ± RSD (%)[a]				
Sample Number	>10 mesh	10-20 mesh	20-40 mesh	40-100 mesh	<100 mesh
35187	0.080 ± 0.006(8%)	0.084 ± 0.007(8%)	0.123 ± 0.005(4%)	0.220 ± 0.015(7%)	0.339 ± 0.009(3%)
35188	1.43 ± 0.28(20%)	1.30 ± 0.17(13%)	2.32 ± 0.17(7%)	2.46 ± 0.13(5%)	2.79 ± 0.08(3%)

[a]Results for 6 runs of 2 samples each

Though the signal increase with decreasing particle size is to be expected, another factor for multilayered paints (*e.g.*, paint 35188) is the differing potential of the various layers to be ground into small particles. Some layers--usually the older ones and the ones more likely to contain lead--are more brittle and flakier, possibly resulting in a higher percentage lead in the fractions containing the finer particles than in those containing the larger particles. This would imply that thorough grinding is important for accurate evaluation of the lead content in the entire sample. Of course, a similar problem could be presumed to be present for acid digestion, in which differing solubilities of layers would also result in erroneous evaluation of the lead content of the paint sample as a whole.

Rather than using a given fraction for analysis, it seemed feasible to simply analyze everything passing a given sieve. Table 4 gives the results for particles passing the designated sieve. Since the results for the 40- to 100-mesh fraction approached those for the fraction passing the 100-mesh, and both had good reproducibility, subsequent studies were done with the fraction passing the 40-mesh sieve. The result for sample 35187 was 0.278% ± 0.011(4%), and that for sample 35188 was 2.71% ± 0.11(4%), obtained from four runs of three samples each, or a total of twelve analyses.

Sample Mass. The next item to investigate was the amount of sample needed in the XRF sample cup in order for the sample to be considered "infinitely thick," that is, of such a thickness that the intensity of the fluorescent radiation is independent of further sample addition. For this investigation, sample 35187 was used in the form of particles passing the 40-mesh sieve. The layer of sample was made uniform in thickness but without expending undue effort to do so. A separate portion was obtained for each run, with the smaller masses being weighed out first. As can be seen in Table 5, only about an 8% decrease was observed in changing from a 0.5-g sample to a 2.5-g sample. Later work, however, indicated that part of this change may have been due to segregation of particles in the fraction smaller than 40 mesh, with the finer ones being toward the bottom of the sample. Thus, the relative amount of the finer particles taken for a given analysis sample would cause the results to vary and to result in higher sample standard deviations. To avoid both this potential for error and the smaller signal observed, analyses hereafter were performed on paint that passed the 100-mesh sieve (see Table 6). Therefore, mass changes above about 1 g were within the method experimental error. The conclusion drawn was that as long as the sample analyzed was not too small (< 0.5 g), concern over weighing the sample accurately does not appear to be necessary.

Table 5. Sample Mass Effects for Paint 35187, Particles Smaller than 40 Mesh

Mass (g)	*Observed Wt% Lead (5 samples)*	*Cumulative Percent Change*
0.3	0.332 ± 0.005[a]	—
0.5	0.317 ± 0.010	-4
1.0	0.299 ± 0.005	-10
1.5	0.301 ± 0[b]	-9
2.0	0.290 ± 0.009	-13
2.5	0.290 ± 0.001[b]	-13

[a]Results for 3 samples [b]Results for 2 samples

Table 6. Sample Mass Effects for Selected Paints, Particles Smaller than 100 Mesh

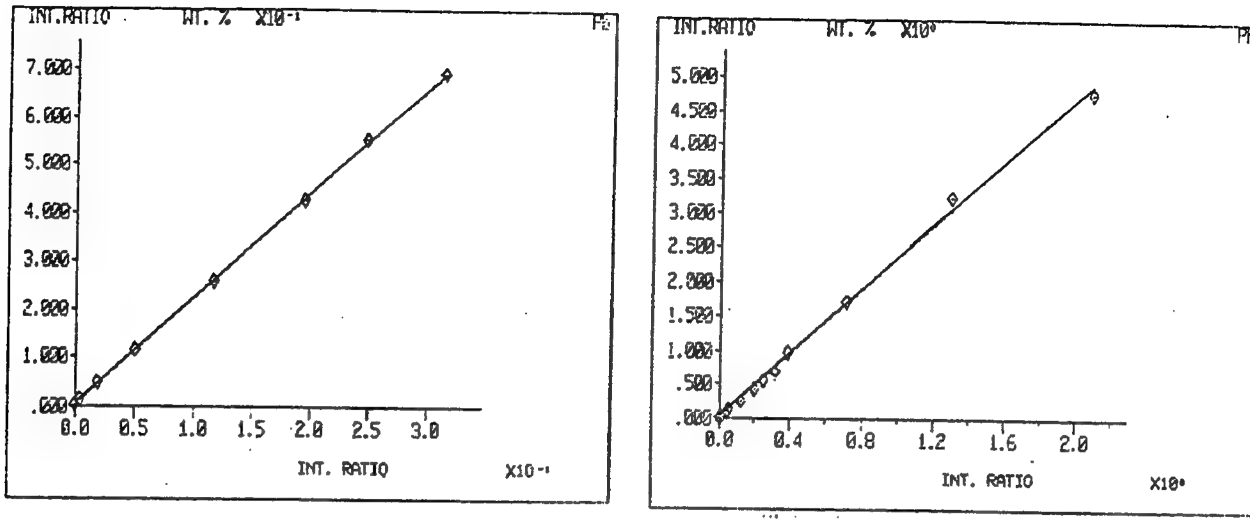
Sample Number	*Mass (g)*	*Average Wt% ± RSD (%)*[a]	*Cumulative Change*
35187	0.7	0.323 ± 0.009 (3%)	—
	1.5	0.293 ± 0.011 (4%)	-9
	2.0	0.287 ± 0.008 (3%)	-11
35188	0.7	2.28 ± 0.12 (5%)	—
	1.5	2.29 ± 0.03 (1%)	<1
	2.0	2.31 ± 0.02 (1%)	+1
35189	0.7	0.725 ± 0.013 (2%)	—
	1.5	0.741 ± 0.022 (3%)	+2
	2.0	0.752 ± 0.012 (2%)	+4
35053	0.7	0.236 ± 0.003 (1%)	—
	1.2	0.228 ± 0.007 (3%)	-3

[a]Results for 3 analyses of each mass

<u>Quantitation</u>

<u>Standards.</u> Twelve paint samples of known concentration were used to obtain a linear plot, using a least-squares (LS) fit, for quantitation. These standards were prepared by Research Triangle Institute (RTI) using real-world paint chips and previously tested by a number of laboratories under the direction of the Environmental Lead Proficiency Analytical Testing (ELPAT) Program, administered by the American Industrial Hygiene Association (AIHA). The "true" values for the standard paints ranged from 0.0481 wt% to 4.77 wt%. The LS plot also contained a "blank" of sodium nitrite (Baker Analyzed, reagent grade) which had a nominal lead concentration of 0.0002 wt%, and an EPA dirt sample with a concentration of 0.012 wt%. An early attempt to obtain a linear plot by least squares exhibited a fair amount of scatter of points at the higher concentrations. However, using equal masses of sample (0.7 g--the limiting sample mass for the standards) in each XRF cup improved the fit. The curve shows excellent fit of all points up through 0.692 wt% (Fig. 1a), with an intercept of 0.0024. Addition of the more concentrated standards (*i.e.*, 0.958, 1.72, 3.22, and 4.77 wt%) to the plot still showed good fit (Fig. 1b) but gave an intercept of 0.0046 and a larger slope (2.34 *vs.* 2.19); root-mean-square (RMS) error was also higher (0.0042

vs.0.00029). One possible answer would be to use two different LS plots: one for higher concentrations and one for lower ones. Due to the difficulty inherent in the programming, there was not sufficient time during this research period to test this, however, and the LS plot containing all points up through 4.77 wt% was used in quantitation.



a. b.

Figure 1. LS Plot for 0-0.69 wt% Lead (a) and 0-4.77 wt% Lead (b)

To test the accuracy of this least squares plot, several other ELPAT paints were analyzed as if they were unknowns, and the observed concentration compared with the true value by reporting the results as percent recovery. The results appear in Table 7.

Table 7. Analysis of ELPAT Paint Samples by Least Squares

Sample Code	*Average Wt% Pb ± RSD (%)*[a]	*"True" Value*	*Percent "Recovery"*
J	0.0484 ± 0.0005 (1%)	0.0481	101
QC-1	0.0168 ± 0.0016 (9%)	0.012	140.
F	0.123 ± 0.004 (3%)	0.114	108
A	0.176 ± 0.013 (7%)	0.201	87.6
B	0.500 ± 0.012 (2%)	0.526	95.0
H	1.60 ± 0.03 (2%)	1.72	93.0
Round 010, #3	3.61 ± 0.04 (1%)	3.58	101
QC-2	3.18 ± 0.10 (3%)	3.22	98.8
Round 011, #1	5.00 ± 0.07 (1%)	4.12	121
Blank	0.0050 ± 0.0004 (8%)	0.0002	—

[a]Results for 4 samples

One possible reason for the high value for QC-1 is that it is a dirt rather than a paint and the particle sizes are significantly larger. All ELPAT paint samples were ground to be ≤ 120 μm. A more probable reason is that the LS plot intercepts the concentration axis above zero, resulting in a calculated concentration for low-lead paints that is higher than it should be. Overall, however, the results are within acceptable ranges.

Detection Limit. There are several methods for reporting detection limits for analytical methods. One such method is to analyze a blank a statistically significant number of times, calculate the standard deviation, and use ten times that value as the detection limit. The blank, sodium nitrite, was analyzed ten times, resulting in an average of 0.0050 ± 0.0005. Hence, on this basis, the detection limit would be 0.005 wt% lead.

Another method for reporting detection limit is the lower limit of detection (LLD)--the amount of analyte that produces a signal large enough so that 99% of the samples containing it will produce a measurable signal. It is calculated by multiplying the standard deviation for a low-level standard by 3.29 (Standard Methods for the Examination of Water and Wastewater, 17th ed., American Public Health Association, p. 1-18). The least concentrated lead standard available in this laboratory contained 0.012 wt% lead. Analyzing this standard ten times resulted in an average value of 0.0150 ± 0.0005. Thus, the LLD by this method is 3.29 times 0.0005 or 0.0016 wt%. Since this value is less than the value obtained for a blank, the first method mentioned is a better measure of the detection limit for the 770 XRF.

Spiked Samples. In order to estimate the method accuracy, samples of paint were spiked with known amounts of lead and analyzed to determine percent recovery of the lead. Paint 35187, ground in a mortar and pestle to pass the 100-mesh sieve, was used for this study. Samples that weighed 1.00 g were

placed into individual 31-mm diameter polyethylene XRF analysis cups that were covered on the bottom with polypropylene. The appropriate amount of calibration standard 1000-ppm lead solution (Environmental Resource Associates, Arvada, CO) was measured directly into the cups with a Hamilton 250-µL syringe; it did not wet the surface of the paint powder but formed powder-coated globules. The cups were then placed in an oven at 95^0 C for one hour, after which time the paint was dumped out on weighing paper, the lumps broken up, and the paint returned to the cup for another hour of drying. The sample was again removed from the cup and gently ground with a mortar and pestle to form a uniform powder. Each sample was finally returned to the original cup, covered with < 4 µm Mylar (polyester) film, and analyzed by XRF. The results, given in Table 8, exhibit excellent method accuracy.

Table 8. Recovery of Lead by Spiking Samples

Lead Added (wt%)	*Volume Standard Solution (µL)*	*Average Wt% ± RSD (%)*[a]	*Expected Wt%*	*Percent Recovery*	*Cumulative Percent Recovery*
0	0	0.234 ± 0.006 (2%)	---	---	---
0.025	250	0.260 ± 0.008 (3%)	0.259	100.4	100.4
0.050	500	0.288 ± 0.006 (2%)	0.284	101.0	101.4
0.075	750	0.312 ± 0.010 (3%)	0.309	99.7	101.0

[a]Results for 3 runs of 3 samples each

<u>Field Samples.</u> With the least squares calibration plot confirmed by paints of known lead concentrations, and with the method accuracy verified, the next step was to analyze field samples of paint and compare the results with the standard method. Some of the paints assayed on the Kevex 770/Analyst 8000 had been previously analyzed in the form of paint chips by the Kevex 700/IXRF system as well as by the standard digestion-AA method; some had been assayed only by AA. Table 9 gives the results of this comparative study. Results are for the analysis of ground paint passing the 100-mesh sieve and for 0.7-g samples, unless otherwise indicated. These sample conditions were used because 1) it had been shown that particle size affected the quantitative results, and 2) the LS plot was obtained using ELPAT paints which had a particle size ≤ 160 µm. Though mass had been demonstrated previously to have little effect on the results, this variable was eliminated by using the same masses for standards and unknowns.

Table 9. Comparison of Methods Using Field Paint Samples

Paint Sample	Average Wt% Pb ± RSD (%) (passing 100-mesh, 3 samples)	Wt% Pb via 700/TXRF (chips)	Wt% via AA (original sample)	Wt% via AA (passing 100- mesh)	Comments
35187	0.323 ± 0.009 (3%)	0.161	0.0762	0.252	≥ 3 layers
35188	2.28 ± 0.12 (5%)	2.44	2.32	—	~ 6 layers
35189	0.725 ± 0.013 (2%)	0.298	0.490	0.657	≥ 4 layers
35053	0.236 ± 0.003 (1%)	0.078	0.058	—	
35048	9.88 ± 0.12 (1%)	4.90	2.34	7.18	
35051	12.26 ± 0.06 (<1%)	20.9	5.64	7.68	
35055	13.16 ± 0.17 (1%)	5.51	0.425	9.36	
35052	0.216 ± 0.009 (4%)	0.258	0.273	—	
35190	0.194 ± 0.003 (2%)	0.071	0.058	—	multilayer
12995	0.114 ± 0.006 (5%)	—	0.0072	—	w/ iron rust
12999	12.94 ± 0.11 (<1%)	—	9.39	—	multilayer
13007	0.309 ± 0.001 (<1%)	—	0.250	—	2 colors
13001	1.38 ± 0.03 (2%)	—	0.96	—	2 colors
12997	9.51 ± 0.15 (2%)	—	8.97	—	
13005	8.62 ± 0.24 (3%)	—	8.39	—	multilayer
13000	0.221 ± 0.004 (2%)	—	0.0680	—	w/ iron rust
13004	0.128 ± 0.005 (4%)	—	< 0.0045	—	w/ iron rust
13006	1.03 ± 0.01 (1%)	—	0.230	—	flexible, multilayer
13009	0.514 ± 0.014 (3%)	—	0.740	—	multilayer, w/ wall
13011	0.010 ± 0.001 (7%)	—	0.0110	—	flexible, w/ cement wall(?)
35056	0.029 ± 0.0004 (1%)	—	0.032	—	w/ caulk(?)
	Mixture of particles and powder				
12996	0.0269 ± 0.003	—	0.0340	—	small amount
13002	0.0957 ± 0.019	—	< 0.0045	—	multilayer, used all
13003	0.106 ± 0.006	—	0.1100	—	substantial amt. of powder
13008	0.0494 ± 0.0136	—	0.0075	—	multilayer, flex., some powder
13010	0.0546 ± 0.0022	—	0.0520	—	some powder

The first part of Table 9 is for paints that could be ground using a mortar and pestle, though a mill would have performed well if it had been available at the time. The portion of the table below the dividing line contains paint samples that could not be ground by mortar and pestle because they were flexible and did not fragment. Even with the mill (Micro-Mill, Model H37252, Bel-Art Products), some paints did not grind well, presumably because their volume was insufficient. (The operating instructions stated that the sample must have a volume between 20 and 50 mL for adequate milling.) Since there was not sufficient powder that passed the 100-mesh sieve in some cases, the entire sample, consisting of small pieces and varying amounts of powder, was analyzed. The powder tended to settle to the bottom of the cup.

It had been noted previously when comparing results for the AA and XRF methods on a set of paints that there were a number of samples that differed greatly. Analysis of the powdered paint by XRF was still greatly different from the results obtained by the AA method. However, reanalyzing by AA the same powdered paint used by XRF resulted in notable changes in the results, bringing them closer to those obtained by XRF on the powder. (See columns 4 and 5 in Table 9, particularly sample 35055.)

The results are sometimes higher by XRF than by AA, possibly for several reasons. One reason may be that the calibration curve for XRF is nonlinear at high lead concentrations. (The most concentrated standard paint available in this lab was 5.06 wt%, so the linearity could not be checked beyond that point.) Another even more likely reason is that, in the digestion-AA method, part of the mass of the weighed sample contains some of the substrate that was painted, making the reported concentration lower than the "true" value. A most important point to be made, however, is that whether analysis is performed by digestion-AA or by XRF, the paint sample *should* be ground!

Though not shown in the Table 9, analysis of these field paint samples was always accompanied by analysis of standard paints which served as a check on the method and on the XRF spectrometer. The standard paints were chosen to bracket as much as possible the expected concentrations, based on previous AA or XRF analysis. Observed concentrations for the standards were always within 10% of the "true" value, except for the 0.012 wt% standard which consistently was found to be about 0.015 wt%. In conclusion, it can be safely said that the results for the analysis of the field samples given in Table 9 are also within 10% of the true value, unless they are in the very low concentration range.

One further point could be made here, and that is that for the limited number of field samples that were compared by XRF using chips and powder, analysis on the powder resulted in only higher values. That is, no false positives or false negatives were generated when using the chips. The AA method, however, did generate a number of false negatives. (See Table 9.)

Observations. In comparing the two XRF systems--the 700/IXRF and the 770/8000--several observations can be made. The first involves the accuracy and sensitivity of the systems. With the present software for the 770, the range that must be used in order to measure the Compton peak for silver (20.3-21.7 keV) is 0-40.92 keV, resulting in a channel width for the multichannel analyzer of 40 eV. However, the analyte peak being used for lead is centered on 12.62 keV, whereas the nearest available channels are

centered on 12.60 and 12.64 keV. On the other hand, the IXRF software allows one to select the range over which to analyze without being constrained to the 10-, 20-, or 40-keV ranges, and results in a channel width of 10 eV. The outcome of this is that one is able to measure the intensity of the lead peak at its maximum, not on one or the other shoulder, resulting in greater sensitivity and accuracy. A smaller standard deviation may also be expected with the narrower channel.

The second observation has been alluded to previously, and that is the ease with which an operator can learn to use the IXRF system, in contrast to the Analyst 8000 FORTRAN programming. Further, the PC-type processing of the IXRF software is more rapid than that of the Analyst 8000, allowing more rapid throughput of samples. The format of the report may also be easily customized to the application desired.

Conclusion. The results of this study do show that XRF analysis of lead in paint is feasible, even at low levels and in difficult matrices, is reasonably rapid, and is comparable or better in accuracy to the presently used acid digestion-AA method. However, the paint sample should be of sufficient quantity to pulverize in a mill and should be analyzed as a powder for best results.

APPENDIX

Sample Preparation. Based on the results of this present investigation, the following procedure is recommended for preparing the sample for most accurate XRF analysis. The sample is placed in the Micro-Mill in pieces small enough to be contained therein. The time necessary to mill the sample depends on its characteristics--i.e., its brittleness or flexibility as well as the amount. About one minute or less will probably suffice for most samples. The contents of the mill are then dumped onto the 100-mesh sieve and the sample passing through it is then placed in a polyethylene XRF analysis cup. If the amount of < 100-mesh sample is insufficient to fill the cup at least 3 mm deep with powder, further milling may be necessary. Inability to obtain sufficient quantity of paint to pass the 100-mesh sieve does not appear to cause large errors for paints that are not multilayered or do not contain high lead concentrations, especially if only a rough estimate of lead content is desired. (See Table 9.) In such cases, simply use the entire milled sample; the finest particles will naturally sift toward the bottom of the cup. Finally, the mill and the sieve should be cleaned with a soft brush or cloth before the next sample is prepared.

Acquisition time for analysis was 100 sec (live time) at 35 keV and 1.3 mA, using the silver secondary target and a 40-eV range.

XRF Analysis. To boot the computer, choose item 5 on the menu for the Program key, then item 5 from the System Setup menu--"Boot DEC Computer." Follow the other prompts as they occur. An asterisk will be displayed when ToolBox™ is running. Once in this software, the programs given below will automatically process a tray of samples containing one to sixteen cups by typing RUN LEAD followed by hitting the RETURN key. The program will then ask for the first and the last tray position numbers. Entry of each number followed by RETURN will then lead to a request for the names of each sample in order. Single integers will be rejected as proper labels, but any other combination of letters and numbers is acceptable. Once entered, the program calls up the processing routine, RUNPNT1, which will take over the spectrum acquisition, processing, and quantitation. All processing is completed on a given sample before the next one is analyzed. Results are printed out on the Okidata Microline 192 printer for each sample individually. Concentrations are given in wt%, but only two decimal places are printed. If more decimal places are needed, each RESULTS file may be examined individually as follows: type EDIT/RES, space, and the file name; then CTRL V twice will get to the page where the results are given to four decimal places. Typing CTRL E exits the editor.

During the RUNPNT1 routine, the software will recall several files by the RECALL command, abbreviated RE. The first is the matrix file, PAINT.1. This file contains the conditions and their associated codes used for analysis, element search lists, element lines to be used in the identification process, and the reference monitor name, as well as other information not relevant to the present application. The second file to be recalled is the least squares file, PNT712, used for the LS quantitation step. Finally, the deconvolution file, PAINT, is recalled. This file contains the information that the peak

to be used in the deconvolution of the XRF spectrum is the Lβ lead peak; it also contains the Compton ratio region (20.3-21.7 keV). Any of these files may be accessed by the EDIT command; for example, typing EDIT/LS PNT712 would open the least squares file for editing. To skip a page, type CTRL V; to back up, type CTRL R. Either of the procedure files may be accessed for editing by typing EDIT/PROC LEAD or EDIT/PROC RUNPNT1. To abort a program while it is running, hit CTRL A.

In essence, the structure of the RUNPNT1 procedure is composed of acquiring a spectrum (ACQ), identifying the analyte (ID/AD) by creating an element list, processing the escape peaks, subtracting the background automatically, deconvoluting by integration and ratioing the intensity to a reference peak, and quantifying by a least squares-ratioing method. Condition 1 is given in line 1 of the procedure and is controlled by the CONTROL command (CON). Windows are cleared (CL) or processed (PRO) as indicated in the procedure. Semicolons in the procedure serve as carriage returns without the need of operator intervention, causing operation to occur on or with the last recalled file.

When finished with the Analyst 8000 for the day, type CTRL C, wait till the XRF ToolBox™ prompt is displayed, push the release button next to the disk, wait for the door to unlock, and remove the disk.

PROCEDURE FILE 0:LEAD:1:0

```
!This procedure establishes starting sample and
!ending sample position, no. of samples and labels
#ist=1
#last=16
!^z
!
!
!~1Place the samples in the sample tray
!
ask/int ' ~1Enter first sample position ' #ist
ask/int ' ~1Enter the last sample position ' #last
!^z
!
! ~1Enter ~3REQUIRED ~1sample labels when prompted
!
#name=' '
$pos=#ist-1
loop $nsamp=#ist,#last
$pos=$pos+1
#name=' '
ask/fil' ~1Enter the sample name for position $pos' #name
setup/names $pos=#name
loop/end
!
pau'~1Press RETURN when samples are loaded and X-rays are~3ON'
!^z
RUN RUNPNT1
```

PROCEDURE FILE 0:RUNPNT1:1:0

```
CON/CON 1     !Air, Ag target, 35 keV, 1.3 mA, 40 keV range
!^Z
!        ~2X-RAY TUBE STABILIZING...
RE/MAT PAINT.1
RE/LS PNT712
RE/DEC PAINT
BA/CL; PRO/WI/CL
CON/POS #ist
LOOP #pos=#ist,#last
!^Z
!        ~2ACQUIRING SPECTRUM...
ACQ/CL/START/WAIT;
ID/AD PB,SE
!^Z
!        ~2PROCESSING SPECTRUM...
PRO/ESC;
BACK/AUTO/SUB;
DEC/INT/RATIO PB LB,-SE;
!^Z
!        ~2CALCULATING CONCENTRATIONS...
QUANT/LS/RATIO/PRINT;;
CON/POS +
BA/CL; PRO/WIN/CL
LOOP/END
!^W
!^Z
!        ~2SETTING X-RAY IN STANDBY MODE...
CON/CON 2
```

COMPUTER SUPPORTED COLLABORATIVE WORK ENVIRONMENT FOR REQUIREMENTS ANALYSIS PROCESSES IN DESIGN OF WEAPON SYSTEMS

Dr. Alexander Bordetsky
Department of Decision Sciences
School of Management

University of Texas at Dallas
P.O.Box 830688
Richardson, TX 75083-0688

Final Report for:
Summer Faculty Research Program
Armstrong Laboratory

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, DC
and
Armstrong Laboratory

September 1995

COMPUTER SUPPORTED COLLABORATIVE WORK ENVIRONMENT FOR REQUIREMENTS ANALYSIS PROCESSES IN DESIGN OF WEAPON SYSTEMS

Dr. Alexander Bordetsky
Department of Decision Sciences
School of Management
University of Texas at Dallas

ABSTRACT

The Requirements Analysis Process in Design for Weapon Systems (RAPID-WS) is a research initiative within the Armstrong Laboratory to improve critical aspects of the DoD weapon systems acquisition process, specifically aimed at improving the requirements process. The proof of concept produced by this research is the RAPID-WS computer system which is characterized by its underlying database management system, graphical user interface, software supporting individual/stand-alone computers, and supporting hardware.

The kernel of existing process for RAPID-WS is represented by Pre-Milestone 0 and Pre-Milestone 1 activities of the DoD acquisition process . These activities include the mission area analysis process as well as the drafting and coordination of certain high-level requirements documents.

Due to the shared electronic document workspace and mixed asynchronous/real-time computer communication features, desktop video conferencing provides the maximal flexibility for distributed group of Action Officers to process all major steps of milestone. It also gives maximum flexibility and highest potential for sharing such applications as a stand alone RAPID-WS, word processor, spreadsheet, modeling, and simulation software.

The subject of this *research* is to study the efficient ways of computer supported collaborative work integration into the Pre-Milestone 0-1 part of the general Acquisition Process. The primary *research tasks* that are addressed in this work include the demonstration of how computer supported collaborative work based on desktop video conferencing technology may be accomplished during the requirements generation process, experimental evaluation of new Pre-Milestone 0-1 related teamwork activities that emerge in computer supported collaborative environment , experimental evaluation of how a defined set of distributed collaborative work tools will support the requirements generation process, experimental evaluation of possibilities to integrate these tools into the RAPID-WS environment, and design of a coordination protocol.

1 INTRODUCTION

The Requirements Analysis Process in Design for Weapon Systems (RAPID-WS) is a research initiative within the Armstrong Laboratory to improve critical aspects of the DoD weapon systems acquisition process, specifically aimed at improving the requirements process. The proof of concept produced by this research is the RAPID-WS computer system which is characterized by its underlying database management system, graphical user interface, software supporting individual/stand-alone computers, and supporting hardware.

Although the stand-alone configuration of RAPID-WS has been received enthusiastically by users, the need for a RAPID-WS environment which supports geographically distributed collaborative work is essential for the purpose of developing RAPID-WS to its full potential as a weapon systems requirement management and analysis tool.

New technologies of high -speed telecommunications enable several collaborative computing environments that are essentially capable of extending collaborative processes of RAPID-WS into a geographically distributed environment by featuring some or all of the following services:

- compressed full motion video,
- shared electronic workspace (shared data models, shared screen, shared graphics and animation),
- high speed file transfer,
- internetworking with switched services.

The forms of communication may vary from asynchronous World Wide Web client-server interactions via Internet to teleconferencing rooms and multipoint desktop video conferencing. The latter incorporates shared workscreen and applications into the "same time-different places" type of distributed electronic meetings. Shared applications include such examples as a stand alone RAPID-WS, word processor, spreadsheet, modeling, or simulation software.

Due to the shared electronic document workspace and mixed asynchronous/real-time computer communication features, desktop video conferencing provides the maximal flexibility for distributed group of Action Officers to process all major steps of milestone [2]. It also gives maximum flexibility and highest potential for sharing such applications as a stand alone RAPID-WS, word processor, spreadsheet, modeling, and simulation software..

The kernel of existing process for RAPID-WS is represented by Pre-Milestone 0 and Pre-Milestone 1 activities of the DoD acquisition process [1]. These activities include the mission area analysis process as well as the drafting and coordination of certain high-level requirements documents.

A key insight into the framework of the Pre-Milestone 0 and Pre-Milestone 1 activities is that the set of activities may be seen as a multidimensional collaborative work space. Each activity may be described as a set of dependencies which must be met before the following activity may begin. Computer supported collaborative technology may effectively facilitate to explore the deficiencies and mission needs statements in the group environment, and helps to share and introduce to the geographically dispersed Action Officers the features of new weapon system solution, including such means as distributed computer simulations, and multimedia demonstrations of new technologies. A coordination protocol [2] will be required for the management of these dependencies in the geographically distributed environment of computer supported collaborative work.

So far, the research in the area of computer supported collaborative work was concentrated on the general communications and user interface issues [3]. In the meantime it's known very little how to integrate the computer supported collaborative work and specifically desktop video conferencing technology into the operational requirements process.

The subject of this research is to study the efficient ways of computer supported collaborative work integration into the Pre-Milestone 0-1 part of the general Acquisition Process.

The primary research tasks that are addressed in this work include:

- the demonstration of how computer supported collaborative work based on desktop video conferencing technology may be accomplished during the requirements generation process,
-experimental evaluation of new pre-milestone 0-1 related teamwork activities that emerge in computer supported collaborative environment ,
-experimental evaluation of how a defined set of distributed collaborative work tools will support the requirements generation process,

-experimental evaluation of possibilities to integrate these tools into the RAPID-WS environment, and
-design of a coordination protocol,

The framework of this research is based on the experimental observations of Air Force Action Officers performing activities of drafting an operational requirements document in support of a documented Air Force need. This is accomplished in a distributed environment using a number of different tool sets to include a desktop video conferencing tool and the RAPID-WS system.

The paper is structured as follows. Section 2 looks at the background information and detailed statements of the research tasks. Section 2.1 describes the pre-milestone 0-1 model for the requirements analysis processes in design of weapon systems. Section 2.2 presents a comparative evaluation of different computer supported collaborative technologies and the reasons why a desktop video conferencing technology was selected for this study. The final part of Section 3 provides the detailed statements for the research tasks based on pre-milestone 0-1 process specifics as well as collaborative features of desktop video conferencing technology. Section 3 describes the method of research. The case for experimental study, search and rescue (SAR) system requirements analysis process, is briefly described in Section 3.1. Communication space of desktop videoconferencing, and configuration of the testbed setup are presented in Section 3.2 and 3.3. Scenarios for experimental tasks are described in Section 3.4. Taxonomy and analysis of experimental results are discussed in the Section 4. They include the requirements to computer supported cooperative work for RAPID -WS (Section 4.1) and ontology of RAPID-WS support agents. Coordination protocol is discussed in Section 5,. The last section (Section 6) summarizes results of the study and directions for further research.

2. BACKGROUND: ACQUISITION MILESTONES AND COLLABORATIVE TECHNOLOGIES

In the background of this research are two emerging technologies. One is a technology of systems analysis for requirements generation processes in design of weapon systems. This technology is based on the Acquisition Milestone and Phases model [3],. It is a complex structure of workflow for acquisition activities and their dependencies. Exit criteria for each milestone must be successfully achieved before a program may enter the next phase of program development. This is a *top-down* ("requirements pull") systems analysis paradigm, that provides a systematic approach for breaking the mission needs statements down to the technology solutions.

The other technology is that of geographically distributed computer supported collaborative work(CSCW). It is a computer technology that is integrated with high-speed telecommunications. Such technology is capable of improving the efficiency of the USAF operating commands teamwork during the different phases of Acquisition Milestone and Phases model. According to [1] one of the major targets for Acquisition Milestones approach is to support the concept of incremental milestones planning. Due to the mixed real-time /asynchronous geographically distributed communications, that collaborative technology is capable to support during the different phases of the entire acquisition process, such technology represents an efficient solution for incremental planning support. Another important target for Acquisition Milestones approach is providing communication channel between the planner, technologist, and user to form the common understanding of the operational need through mission area analysis. It means collaborative work on mission area analysis documents from geographically distributed locations. This task may be naturally addressed by computer supported collaborative work technology.

The advanced forms of CSCW technology may provide solid and immediate contribution to the improvement of Acquisition Process.

2.1 COLLABORATIVE FEATURES OF PRE-MILESTONE 0-1 MISSION NEEDS STATEMENT ANALYSIS PROCESS

According to [3] the most critical phase of the entire Acquisition Process is that of Pre-Milestone 0-1 activities. During this critical phase, the important projections, assumptions, and decisions are made impacting the success or failure of an emerging system program. Figure 1 represents the traditional workflow diagram for Pre-Milestone 0-1 activities. As a part of general acquisition milestone analysis model Pre-Milestone 0-1 model represents a *top-down* ("requirements pull") design paradigm,

that provides a systematic approach for breaking the mission needs statements down to the technology solutions. It is a problem oriented approach that starts from deficiencies analysis and access the technologies that require improvement, or do not even exist. In many practical Air Force acquisition cases [1] the laboratory or industry sponsor presents the technology that significantly increases capability. The solution is well understood , and user needs to identify the requirements. This represent a *buttom-up,* or "technology push" paradigm in the requirements generation process. If the general Acquisition Milestone Phases model and Pre-milestone 0-1 process implement the *top down* paradigm, collaborative activities at many steps of Pre-Milestone 0-1 process may effectively combine both approaches. Computer supported collaborative technology may effectively facilitate both approaches: it helps to explore the deficiencies and mission needs statements in the group environment, and helps to share and introduce to the geographically dispersed Action Officers the features of new weapon system solution, including such means as distributed computer simulations, and multimedia demonstrations of new technologies.

Most of the steps in Pre-Milestone 0-1 process are collaborative by their nature. Some of them are processed only once during the Pre-Milestone 0-1 process (Fig 2) , the others repeat many times. The taxonomy of Pre-Milestone 0-1 collaborative processes may be described as follows:

-Group decision making sessions in support of the ***Mission Area Assessment (MAA)*** and ***Mission Needs Analysis (MNA).*** The MNA is where services asses their capability to complete the operational tasks under assigned mission area. The MNA provides the opportunity to analyze mission needs with respect to the effectiveness of existing systems against the specific mission on a task list. Both of these activities are accomplished on a recurring basis to provide the necessary detail to the mission area plan.

-Collaborative (multiperson) ***production of the document.*** This activity repeats many times for different documents such as Mission Area Plan (MAP), Mission Needs Statement (MNS), Cost and Operational Effectiveness Analysis (COEA), and Operational Requirements Document (ORD).

-Group ***review of the documents.*** The activity repeats many times for such steps as reviewing the MNS and ORD content and format, review of the documents by the Air Force Requirements Oversight Council (AFROC) and the Joint Requirements Oversight Council (JROC, and COEA study plan development review.

-Group ***review and prioritization of alternatives.*** This activity is required for integrating and prioritizing of MAP by the joint team of Action Officers and senior personnel. It is later applied in the selection of alternatives from the COEA or other studies in the review and selection of alternatives that form the basis for the ORD.

-***Conflict resolution*** activity is used to support collaborative work during the reconciliation of comments related to the draft of MNS, and the draft of ORD.

-***Routing and distribution*** activity is used to support the MNS and ORD document review and signing process during the "for comment" phase and Air Staff coordination.

Most of the listed activities presume the mixed real-time (face-to-face) and asynchronous group communication sessions for their performance. The Air Staff involved in the process is initially geographically distributed, and each step of Pre-Milestone 0-1 process is rather time consuming. Computer supported collaborative technology facilitates the problems of geographically distributed group work. It represents an advanced resource for saving the acquisition process time and brings a new quality to the requirements generation process automation.

2.2 COMPUTER SUPPORTED COLLABORATIVE TECHNOLOGIES BASED ON HIGH-SPEED TELECOMMUNICATIONS

The forms of computer supported collaborative work may vary [4] from asynchronous World Wide Web client-server interactions via Internet to teleconferencing rooms and multipoint desktop video conferencing. The latter incorporates shared workscreen and applications into the "same time-different places" type of distributed electronic meetings. Shared applications include such examples as a stand alone RAPID-WS, word processor, spreadsheet, modeling, or simulation software. Due to the shared electronic workspace and mixed asynchronous/real-time communications desktop video conferencing provides the maximal potential for incorporating decision support models into the distributed collaborative work.

2.3 DESKTOP VIDEO CONFERENCING

Desktop video conferencing is the central representation for a collaborative computing environment. It makes available at the working desk all types of services, required for distributed multimedia collaboration:

- personal video conferencing (compressed full motion video),
- shared electronic workspace (shared data models, shared screen, shared graphics and animation),
- high speed file transfer,
- internetworking with switched-based services.

An advantage of desktop video conferencing is that members are located in their own offices where each person has access to his or her own resources; however they also have distractions: phone calls, e-mail arrivals, and visitors [5]. Video conferencing rooms allowed only scheduled meetings in the rooms isolated from outside interrupts; but unlike video conferencing rooms desktop conferencing allows spontaneous interactions between individuals or small groups.

In the desktop video conferencing environment decision makers (DMs) communicate through desktop computers. The computers are equipped with video cameras, compression boards, and different adapters that link them to local and wide area high-speed switched telecommunications networks. Basic rate Integrated Digital Services Network (ISDN) provides the lowest level of desktop video conferencing communication. Fiber optic and satellite channels, connected to local computers via Asynchronous Transfer Mode (ATM) switches provide the highest level of wide area communication respectively. Decision makers may be located in the same room, or in the same building. They may distributed campus wide , or may be dispersed geographically , country and world wide.

No matter what level of reach in communication is used the collaborative work in the desktop video conferencing environment is initiated by means of voice and video calls. typically the voice call goes first. Then video call provided by such feature as compressed full-motion video allows two or more parties to see each other , establish in-person communication , and start up a meeting. Several options may be used for sharing the documents such as MNS, COEA, and ORD, collaborating on comments, and brainstorming about the document corrections. The minimal set includes the shared workscreen, high-speed file transfer, and a version of chat box. All tools may be used concurrently to video/voice communication.

One distinctive feature of desktop video conferencing is that none of the participants in a meeting need to be physically located in one room. They are distributed through their workplace and communicate via personal computers. One of the first experimental studies of multiparty desktop conferencing via ISDN lines was done with the MERMAID system [7]. The MERMAID system has been successfully tested in meetings connecting up to 4 locations, with the number of participants ranging from 2 to 8. The topics of these conferences include software specifications design, planning of research and development activities, and some other system analysis and design issues. Results are the following:

- Participants have noticed little delay in data transmission, even when number of them has increased up to more than four,
- Voice delay was too negligible to be noticed by participants. Although the video signal delays ranged between 0.2 to 0.5 seconds on ISDN (128 Kbps), participants rarely become impatient. This may be due to the capability of transmitting the facial expression of a participant in less than 0.3 second.
- Participants noticed delay in the transmission of handwritten data and window manipulations. Some became impatient, because it takes 10 to 20 seconds to send and display shared document.
- When more then four participants joined the conference there has been some difficulties in determining who was speaking,
- Superiors and subordinates favored the designation mode to invoke shared windows,
- Persons of nearly equal rank preferred the first-come-first-served mode,

- In brainstorming sessions free mode was used.

Researches from Sun Microsystems [5] have studied 72 desktop video conferences between two remote locations in Massachusetts and California. Their observations revealed the following details:

- Group interaction that occurred in desktop video conferencing was more like in face-to-face meetings. Remote collaborators were able to interrupt each other, accomplish turn completion, and even occasionally joke (features that were markedly absent in video conferencing room),
- During the study, one-way audio delays were measured between 0.22-0.44 second, which is better then 0.57 seconds delay in the video conferencing room, but still noticeable,
- Desktop video conferencing reduced significantly the number of e-mail messages per day,
- Desktop video conferencing almost eliminated the use of video conferencing rooms.

Comparing the previous observations made on both video conferencing rooms and desktop video conferencing technologies, we may conclude in general, that some serious communication constraints exhibited in video conferencing rooms are not observed with desktop video conferencing. Moreover, desktop video conferencing demonstrated less sensitivity to voice delays, immediate access to shared data model, and ability to communicate more like in face-to-face meeting.

The next section describes the approach to experimental study of the efficient ways of linking desktop video conferencing technology to Pre-Milestone 0-1 part of Acquisition Process for RAPID-WS.

3. METHOD OF RESEARCH

The method of this research is basically experimental. It is based the observations of users ability to perform the RAPID-WS process in the physical environment of desktop video conferencing testbed. The following three approaches was used to conduct the observations:
-playscripts of Action Officers behavior during the experiments,
-video records of collaborative work in front of the desktop, and
-questionnaires

Search And Rescue (SAR) system operational requirements design process was selected as a case study for the experiments. The experiments was structured to test the performance of basic Pre-Milestone 0-1 collaborative activities (Section 2.1) in the range of teamwork support tools that may be integrated under desktop video conferencing environment

3.1 THE CASE FOR EXPERIMENTAL STUDY: SEARCH AND RESCUE SYSTEM REQUIREMENTS GENERATION PROCESS

Search And Rescue (SAR) communication system is a complex of three major telecommunication segments: the user segment consisting of the hand-held survival radio, the space segment allowing over the horizon communications, the command control, and communications segment consisting of hubs, base stations, and the rescue aircraft The space segment consists of communication transponders on satellites, or high altitude aircraft. It offers the most significant SAR technical and programmatic challenge of integrating SAR functions with the satellite. The user segment presents challenges in miniaturization. The C3 segment has the technical challenge of making the base station a compact suitcase size computer and display that may which must be aircraft and ship mobile.

The basic structure for SAR Analysis process may be represented through the following phases:
-*Mission needs analysis phase*. During this phase mission needs statements are systematically exploded down to the functional objectives, constraints, and operational scenarios. Mission needs analysis phase is a major top-down analysis process component, that implements milestone approach.
-*SAR alternative generation phase*. At this phase the performance objectives, constraints, and operational concepts derived from the Mission Needs Statements are first time linked to the available SAR systems alternatives. In terms of top-down and buttom-up design approaches, this phase provides the integration of two approaches.

-SAR alternative analysis. This is the most computentially intensieve phase. The measures of effectiveness generated earliar are now applied to provide the trade-off analysis and selection of alternatives.

The tasks of this experimental study were concentrated mostly on the issues specific for the two phases of SAR analysis process: mission needs analysis phase, and SAR alternative analysis.

3.2 CONFIGURATION OF EXPERIMENTAL SETUP

In the desktop video conferencing environment DMs communicate through desktop computers. The computers are equipped with video cameras, compression boards, and different adapters that link them to local and wide area high-speed switched telecommunications networks. Basic rate Integrated Digital Services Network (ISDN) provides the lowest level of desktop video conferencing communication. Fiber optic and satellite channels, connected to local computers via Asynchronous Transfer Mode (ATM) switches provide the highest level of wide area communication respectively. Decision makers may be located in the same room, or in the same building. They may be distributed campus wide , or geographically by state, or even country. In order to simulate this environment the following testbed configuration was used.

During each experiment the collaborative work in the desktop video conferencing environment was initiated by means of voice and video calls. typically the voice call was used first . Then video call provided by such feature as compressed full-motion video, was used to allow two parties to see each other , establish in-person communication , and begin a meeting. Several options were available for sharing the RAPID-WS work. The minimal set included the shared workscreen, high-speed file transfer, and a version of chat box. All tools were be used concurrently to video/voice communication.

Two PCs were equipped with Northern Telecom desktop video conferencing units, and were allocated in two rooms separated by the hall. Both nodes were connected by ISDN network through the UTD-based Meridian PBX switch. Connectivity was provided by Tel Adapters [8] that mixed data, voice, and video channels and transferred mixed packets through ISDN line.

3.3 STRUCTURE AND DESCRIPTION OF EXPERIMENTAL TASKS

The experiments are structured so, that to make possible to study the performance of similar Pre-Milestone 0-1 activities in increasingly more sophisticated collaborative work support environment. In the first experiment only MS Word is used as an application software on the top ISDN-based desktop videoconferencing features of Visit Video. In the second experiment the capability to mix Internet and ISDN communication is added through the WWW features. It is to check how the group communicating through the desktop video conferencing may at the same time access the remote data stores available through the other global network. The third experiment additionally includes RAPID-WS system as a Pre-Milestone 0-1 domain-oriented application, and also includes LAN connectivity under the coordination protocol simulated through the Vision Quest group support. In the third experiment we have three networks working simultaneously: local area network, Internet, and ISDN. We also have a prototype of the coordination protocol delivered by LAN-based Vision Quest application. And we can experimentally evaluate the value of domain oriented RAPID-WS system for decision making in such integrated collaborative environment. The detailed description of three major experiments is as follows.

<u>Experiment #1.</u>
Scenario: general practice with communication system, analysis of current SAR system shortfall, structuring and selecting of system boundary conditions.
Software/communications configuration:

-basic components of Northern Telecom desktop video conferencing system Visit Video,
-ISDN network,
-MS Word.

Figure 3 illustrates the exact sequence of tasks for the first experiment.

Experiment #2.
Scenario:
-including WWW pages into general practice with communication system,
-providing analysis of the current SAR system shortfall,
-deconflicting two documents from SAR COEA,
-prioritizing of 16 SAR system alternatives, selecting the most appropriate one.
Software/communications configuration:
-basic components of Visit Video,
-ISDN network,
-MS Word,
-WWW browser (Netscape).
Figure 4 illustrates the sequence of tasks for experiment #2.

Experiment #3.
Scenario:
-general practice with the desktop video conferencing system,
-analysis of current SAR system shortfalls,
-structuring and selecting the system boundary conditions.
Software/communications configuration:
-basic components of Visit Video,
-ISDN network,
-LAN network,
-RAPID-WS
-Vision Quest,
-Netscape WWW Browser
-MS Word,
Figure 5 illustrates the sequence of tasks for experiment #3

4. ANALYSIS OF EXPERIMENTS AND DESIGN SOLUTIONS

We describe the results of experiments as taking place in the multidimensional space of desktop video conferencing. communication space. The relationships among tested Pre-Milestone 0-1 activities and collaborative work support functions that were identified in result of experiments are presented mainly in the table-like form. Multiple tables represent the results of playscripts, questionnaires, and video records analysis.

4.1 COMMUNICATION SPACE OF DESKTOP VIDEO CONFERENCING

We consider the teamwork in the distributed environment of ISDN-based desktop video conferencing system as a continuing process of Pre-Milestone 0-1 activities developing through the synchronous and asyncronous group sessions on the time scale. Desktop video conferencing technology is capable of support such process along the following dimensions of team collaborative functionality:
-Generic steps of the electronic meeting cycle
-Reach: the level of geographical distribution for DMs,
-Range: the level and types of information sharing,
-Responsiveness: the distribution of syncronous and asyncronous team conferences,
-Coordination: partitioning of management functions, roles of decision makers.

Conceptually, at each phase of SAR systems requirements analysis a full-scale electronic meeting is taken place, but the distribution of collaborative activities related to generic electronic meeting steps would be obviously different at each phase of SAR requirements generation. The generic steps for centralized electronic meeting typically incorporate the following teamwork activities (the alternative may be a model of multiple management roles):

1. Facilitator briefly introduces the agenda : issues to address, number of questions, ability to share the comments, and how the group may come up to consensus.

2. Group tries to adjust the environment by starting verbally to discuss some critical items. This is the first look at the problem. Group organizes the ideas.

3. On the basis of the group discussion, the participants start their individual work on commenting the agenda items. This is a brainstorming part of a session.

4. The participants work with facilitator to understand and consolidate the items in brainstorming lists.

5. Participants vote on the items in the consolidated list. This is essentially anonymous process of client-server interactions with decision support system. Implementation of discrete multiple criteria models is highly desirable at this phase in order to speed up this phase and make it more analytical.

6. As the alternatives discussed over meeting agenda are selected in Phase 5, the group may start the resource allocation problem solving for selected alternatives. This may take several hours, or become a reason for the next meeting. Usually quantitative models become critically important for support of this phase.

We may assume that at the beginning of SAR requirements analysis process, at Mission Needs Analysis phase, the most of the time is spent for generic steps 1, 2, and 3. On the other hand for the SAR alternative evaluation phase we may expect most of the meeting time to be spent for selecting of preferred solutions by using such means as trade-off analysis, analytical hierarchical processes, etc.

4.1 PLAYSCRIPT ANALYSIS

The results of playscript analysis are used as a source for answering the questions related to the such research tasks as

-experimental evaluation of new Pre-Milestone 0-1 related teamwork activities that emerge in computer supported collaborative environment ,

-experimental evaluation of how a defined set of distributed collaborative work tools will support the requirements generation process,

-experimental evaluation of possibilities to integrate these tools into the RAPID-WS environment, and

-design of a coordination protocol.

The results were grouped as follows. For each group we have a set of tables that describe the desktop video conferencing aided performance for a set of RAPID-WS activities related to the task. In the table each activity (or the group of them) is associated with the obtained time for it's performance, and help that was requested or desired by the subjects. An example of such table for the experiment with the intensive usage of different desktop video conferencing features is represented in Table 1.

4.2 QUESTIONNAIRE ANALYSIS

The main objectives for this part of the study is to find out the users views on the value of such features as user interface, general utility, and capability to improve RAPID-WS process that are delivered by desktop video conferencing environment. Three major scales were used to evaluate the user answers: nominal scale, ordinal scale, and interval scale. Any time the interval scale was applied the following format was used:

N/A	0	1	2	3	4	5
Not Applicable	Strongly Disagree	Disagree	Somewhat Disagree	Somewhat Agree	Agree	Strongly Agree

In order to collect the views that would cover the range of potential users the questionnaires were applied to the following categories of users:

-Generally experienced Action Officer,

-Acquisition expert,

-Advanced computer experienced Action Officer,

-SAR experienced Action Officer

4.2.1 USER INTERFACE

The answers about user interface, utility evaluation, and process improvement were grouped in the set of tables such as in Fig.2. The most informative user answers are as follows.

Comment #1. I think the auto phone functions (not really used here) in the VISIT VOICE module would be useful daily, providing speed-dialing and hands-free telephone communication. The file sharing would be useful often, but even more useful-I'd--imagine would be the ability to share applications. The screen share function was very useful, but kluge-ie. It would be easier to share Word then cu from Word , paste into Screen Share, and then paste back into Word

4.2.1 UTILITY EVALUATION

Comment #1. Ability to share applications. Get rid of all but application sharing, video conferencing, and fast file transfer, and you've got a winner. may be add ability to overlay post-it type notes on application sharing.

Comment #2. Automatic sizing and location for open windows might help. Too much time is spent on moving and re-sizing windows. Better edit capability in Screen share. The ability to share more then one screen of data at one time, if possible. As it is Screen Share in it's utility to be able to share only one screen of data at one time. Overall, Visit appears to be pushing the limits on MS-DOS and Windows as operating system environment. Also more capable hardware (CPU, speed, memory, and monitors) will be need to make VISIT really beneficial.

Comment #3. APPLICATION SHARING

Comment #4. Better word processing capability in share screen. Need maximum screen size.

4.2.3 PROCESS IMPROVEMENT

*Comment #1.*__We definitely need to move to a more paperless and integrated (data and processes) office environment, and Visit is a step in the right direction. In the future, I can see on-line availability of automated information as satisfying virtually all our information needs. All information would be accessible through the desktop terminal. Tools, like Visit, definitely are needed to further automate the workplace. I'm just not sure that Visit offers a lot of utility, at this time (needs more development).

Comment #2. Possibly quicker development and coordination of requirements

Comment #3. I see Visit as being a very important step in the right direction, but of relatively limited value at this time. As with any new capability, it will probably improve over time. Probably the biggest limitation with Visit is the hardware and software constraints presented by current microcomputer systems. MS-DOS and Microsoft Windows 3.1 appear to be pretty well mixed out to run Visit. Also, much more capable computers and video (larger monitors, better resolution) are needed to realize the potential of Visit.

Comment #4. In general the Visit software system was quite easy to use. Some of the activities are a bit involved, but not over-burdensome considering the benefits derived - consider the time it might take to ftp the same file instead of calling someone and saying I'm going to send you a file, click, click, click...here it comes. I'm still not convinced of the utility of the video...

4.3 SUMMARY OF OBSERVATIONS AND THE REQUIREMENTS TO CSCW ENVIRONMENT FOR RAPID-WS

According to the dimensions of team collaborative functionality all the experiments had the same level of reach: an ISDN wide-area connectivity between the same two testbed rooms. They differ though in the level of range. If the first experiment had only MS Word program added to the generic desktop video conferencing software, the third experiment incorporated RAPID-WS, WWW, and LAN-based group support system on the top of the desktop video conferencing system.

Inspite the different background and managerial roles all of the teams participating in the experiments spontaneously found out one and the same two-role coordination pattern. This pattern may be described by splitting the leadership for the software control and problem analysis: one site takes a lead

in monitoring software environment, while the other is guiding the document analysis and problem solving.

User interface requirements: All questionnaires and playscripts indicate the same basic interface support requirement. It is automatic structuring of multiple shared windows on desktop video conferencing screen according to the specific collaborative task, user personal preferences, and screen limitations at each site. We will identify the appropriate process as structuring agent.

Relating MNS tasks to deficiencies. One of the major requirements for support of this category of collaborative activities is to transfer the transfer of knowledge about the prototypes. we will identify the appropriate process as a memory agent.

Collaborative production and group review of the documents. For this category of collaborative activities one of the major requirements is to provide the guidance for both sites in MNS, COEA, ORD, and other basic documents analysis We will identify the appropriate processes as the guiding agents. Stand alone RAPID-WS is a good example of such agent. Another typical requirement is to support multiple transactions with basic document sharing tools: screen share, file transfer utility, and application program. Among such functions are saving the changes from the shared screen to the MS Word document, renaming the files for high-speed file transfer, etc. We will identify the appropriate process as application sharing agent.

Review and prioritization of alternatives. The major requirement to this category is that of support for decision matrix generation, editing, ranking, and ordering the alternatives. We will identify the appropriate process as a group decision support agent.

Conflict resolution activities require primarily the same form of support that is for collaborative production and group review of the documents. It is the process of guiding agent.

Coordination support requirements are based on the following conclusions. The coordination patterns are not well studied yet, and it seems to be could hardly be learned in advance. They depend on the number of Action Officers involved one group session, their managerial roles, and their geographical distribution. In the proposed approach it assumed that coordination will be provided by people communicating via desktop video conferencing. The hierarchy of support agents is assumed to facilitate the coordination process among the others. Since the coordination patterns can not be acquired in advance with the sufficient level of detail, the evolutionary approach, that would allow the groups to accumulate the experience from the previously conducted session, and transfer over time and space of the Pre-Milestone 0-1 process is highly desirable. this in turn would require the learning component, group memory agents, to be incorporated in the support environment.

4.5 SOLUTIONS: ONTOLOGY OF THE RAPID-WS SUPPORT AGENTS

The architecture of the proposed collaborative agent-based mechanism is represented by three components: a learning component , a rule-based component, and a collaborative component.

The learning component is responsible for classification and learning, and provides *group memory* features. It is based on case-based reasoning and artificial neural network agents. The case-based reasoning component represents an analogical type of reasoning and makes it possible to create a case-memory for management. It also enables automated capture and maintenance of group knowledge about fault and configuration management. Distributed experts provide the initial knowledge base for likely cases by populating (in natural language) an appropriate set of frames. The case-based reasoning technique then makes it possible to compare a current situation with a set of known solutions and rank it for match. This is required to support the tasks where a formal set of rules is unavailable or difficult to obtain or interpret, but for which examples of correct solutions are readily available. Previous solutions are stored as a case memory. When a search fails to locate a similar case, the search itself becomes the basis for creating a new case.

Artificial neural network is a perfect tool for capturing and transferring the quantitative knowledge for tasks such as SAR boundary conditions analysis. As a group memory agent, it runs concurrently to real-time and asynchronous sessions that are conducted under the desktop video conferencing environment

The rule-based agents are responsible for support of well-structured decisions. These occur more frequently at the initial stages of upper level activities such as general instructions on Preparing MNS (review and staffing, validation and approval), Mission and Threat Analysis, Identifying Materiel and Nonmateriel alternatives, Identifying multiple constraints, etc. Screen structuring agents represent another category of rule-based agents that support a collaborative interface. Rule-based agents may be located on geographically dispersed nodes. Case-based group memory supports their coordination by using a synchronization model (Show and Fox 1993) showing the links to rules at other nodes affected by case-related actions.

The collaborative component provides upper level (manual and knowledge-based) coordination by means of real-time/asynchronous communication based on shared electronic workspace, compressed full-motion video, and high-speed file transfer. It is an evolution of multi-person desktop video conferencing technology. Integrating case-memory , neural net memory, and rule-based agents into a shared electronic workspace supported by "same time-different place" electronic meetings results in a RAPID-WS decision support system that is able to evolve into a hybrid human-machine knowledge-based structure

5. COORDINATION PROTOCOL

Coordination protocol represents the specifications for coordinated objects (processes) and dependency models that are capable of coordinating RAPID -WS processes in desktop video conferencing distributed environment. Models are strongly linked to behavioral roles of collaborating Action Officers.

Basically the coordination protocol provides the mappings among the major dynamic processes:
-collaborative functions for Pre-Milestone 0-1 process,
-the sequence of steps of distributed electronic meeting,
-communication process via desktop video conferencing technology.

For the different Pre-Milestone 0-1 collaborative functions the concentration on different steps of distributed electronic meeting is different. Figure 6 illustrates the example of coordination protocol designed by means of the group support system Vision Quest. In this example transactions with different agents are coordinated by means control structure for the steps of distributed electronic meeting that is coded in Vision Quest. This prototype solves the simples level of coordination-automating the links to the different support agents for support of the SAR alternatives evaluation.

The study of multiple forms of coordination protocols emerging in multiperson meetings is the subject of future research.

5. CONCLUSION AND FUTURE RESEARCH DIRECTIONS.

The primary research tasks that are addressed in this work include the demonstration of how computer supported collaborative work based on desktop video conferencing technology may be accomplished during the requirements generation process, experimental evaluation of new pre-milestone 0-1 related teamwork activities that emerge in computer supported collaborative environment , experimental evaluation of how a defined set of distributed collaborative work tools will support the requirements generation process, experimental evaluation of possibilities to integrate these tools into the RAPID-WS environment, and design of a coordination protocol.

The results of described experiments demonstrate high potential of the desktop video conferencing technology to effectively support the distributed RAPID-WS processes. The results enabled to identify the requirements to the basic support process for conducting the RAPID-WS analysis in the desktop video conferencing environment. On the basis of the identified requirements the architecture of intelligent support agents for RAPID-WS processes was proposed.

The major directions for future research include prototyping of the proposed agent-based architecture, study of multiperson desktop video conferencing functionality, and proof of concept experiments with high-speed B-ISDN, ATM, and advanced satellite communication technologies.

ACKNOWLEDGMENTS

The author would like to thank Capt. Frank Simcox for his great contribution to this work , Janet Peasant for her help and support, and Capt. Ken Moen for helpful discussions.

REFERENCES

[1] T.R. Furguson, Lt. Gen., USAF and T.R. Hertz, Requirements Planning, *Airpower Journal,* Summer, pp.3-17, 1990

[2] Mission Needs and Operational Requirements: Guidance and Procedures. Air Force Instruction 10-601, 31 May, 1994.

[3] T. Malone and K. Crowston, The Interdisciplinary Study of Coordination, *ACM Computing Surveys,* Vol. 6, No. 1, pp. 87-119, 1994.

[4] D. Coleman and R. Khanna, Eds., Groupware: Technology and Applications Prentice Hall, 1995.

[5] J.Tang and E. Isaacs, Why Do Users Like Video? Studies of Multimedia -Supported Collaboration. *Computer Supported Cooperative Work,* International Journal, Vol.1, No.3, pp. 162-196, 1993.

[6] S.Gale,. Desktop Video Conferencing : Technical Aspects and Evaluation Issues. *Computer Communications,* Vol. 15, No.2, pp.517-526, 1992

[7] Watabe, K. et al. Distributed Multiparty Desktop Conferencing System: MERMAID. *CSCW Proceedings*, Oct., pp. 27-38.,1990.

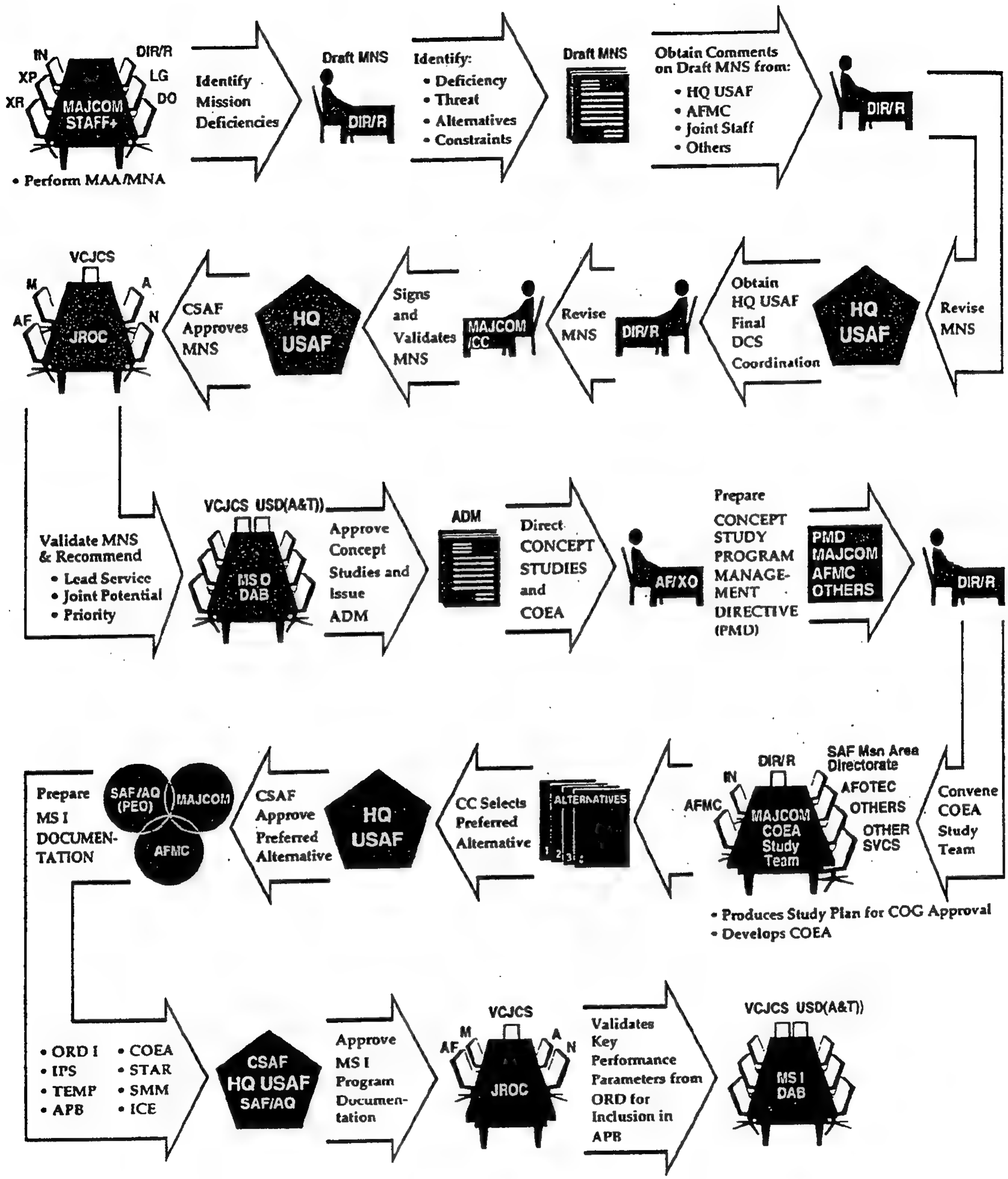

IN
DIR/R
XP
LG
XR
DO
MAJCOM STAFF+
• Perform MAA/MNA
Identify Mission Deficiencies
Draft MNS
DIR/R
Identify:
• Deficiency
• Threat
• Alternatives
• Constraints
Draft MNS
Obtain Comments on Draft MNS from:
• HQ USAF
• AFMC
• Joint Staff
• Others
DIR/R
Revise MNS
HQ USAF
Obtain HQ USAF Final DCS Coordination
DIR/R
Revise MNS
MAJCOM /CC
Signs and Validates MNS
HQ USAF
CSAF Approves MNS
VCJCS
M
A
AF
N
JROC
Validate MNS & Recommend
• Lead Service
• Joint Potential
• Priority
VCJCS USD(A&T))
MS 0 DAB
Approve Concept Studies and Issue ADM
ADM
Direct CONCEPT STUDIES and COEA
AF/XO
Prepare CONCEPT STUDY PROGRAM MANAGEMENT DIRECTIVE (PMD)
PMD MAJCOM AFMC OTHERS
DIR/R
Convene COEA Study Team
DIR/R
SAF Msn Area Directorate
IN
AFOTEC
AFMC
OTHERS
OTHER SVCS
MAJCOM COEA Study Team
• Produces Study Plan for COG Approval
• Develops COEA
ALTERNATIVES
CC Selects Preferred Alternative
HQ USAF
CSAF Approve Preferred Alternative
SAF/AQ (PEO)
MAJCOM
AFMC
Prepare MS I DOCUMENTATION
• ORD I
• IPS
• TEMP
• APB
• COEA
• STAR
• SMM
• ICE
CSAF HQ USAF SAF/AQ
Approve MS I Program Documentation
VCJCS
M
A
AF
N
JROC
Validates Key Performance Parameters from ORD for Inclusion in APB
VCJCS USD(A&T))
MS I DAB

Fig. 1

RRG - Requirements Review Group (Action Officers)
RRB - Requirements Review Board (Colonels)

TRADITIONAL PRE-MILESTONE 0 MNS DEVELPMENT PROCESS

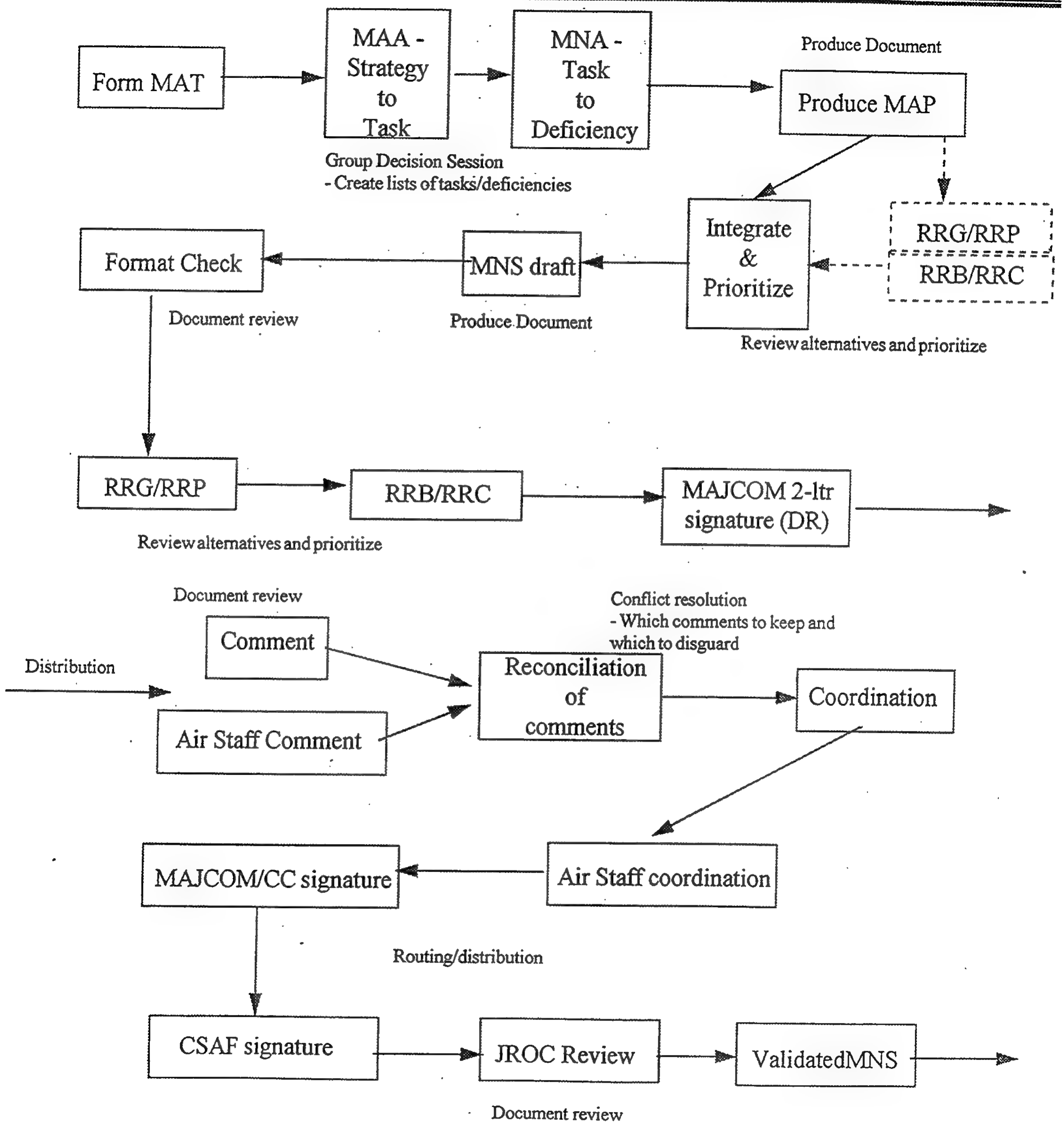


Fig. 2

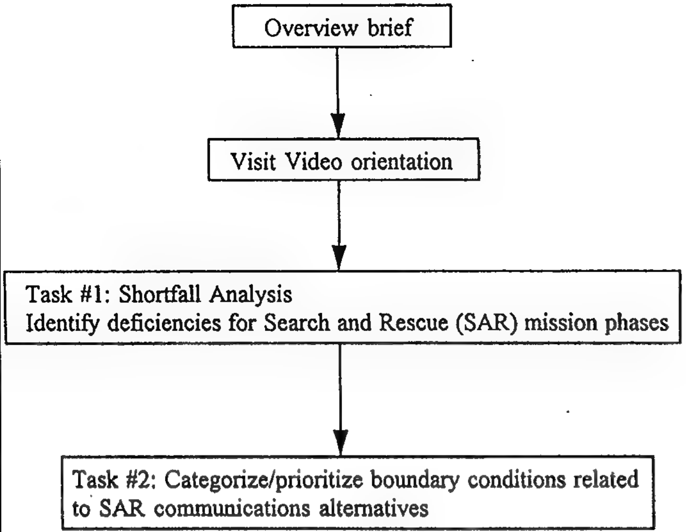


Fig. 3

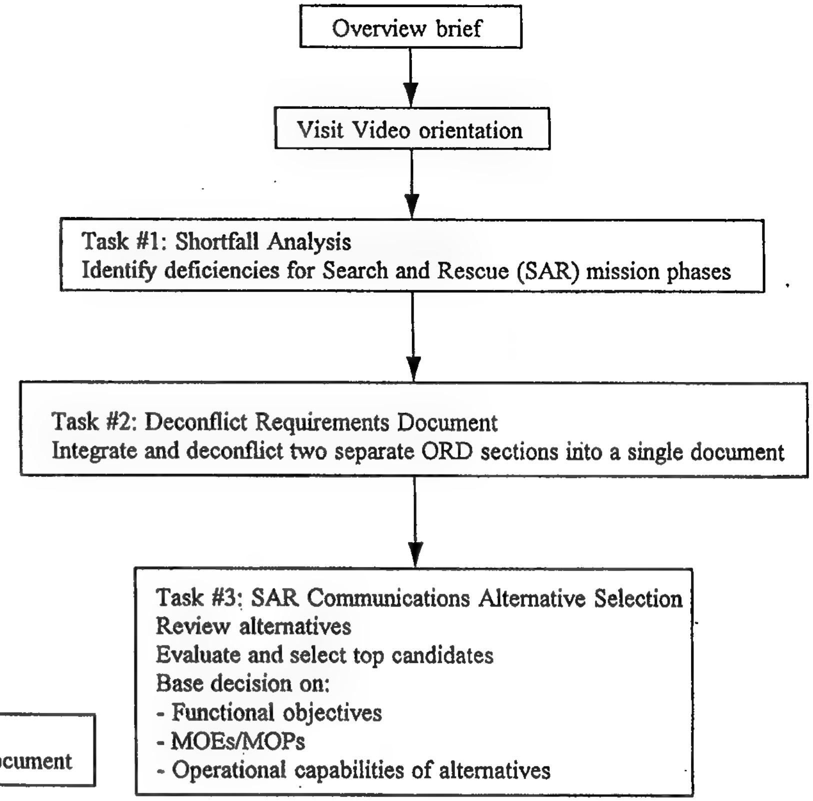


Fig. 4

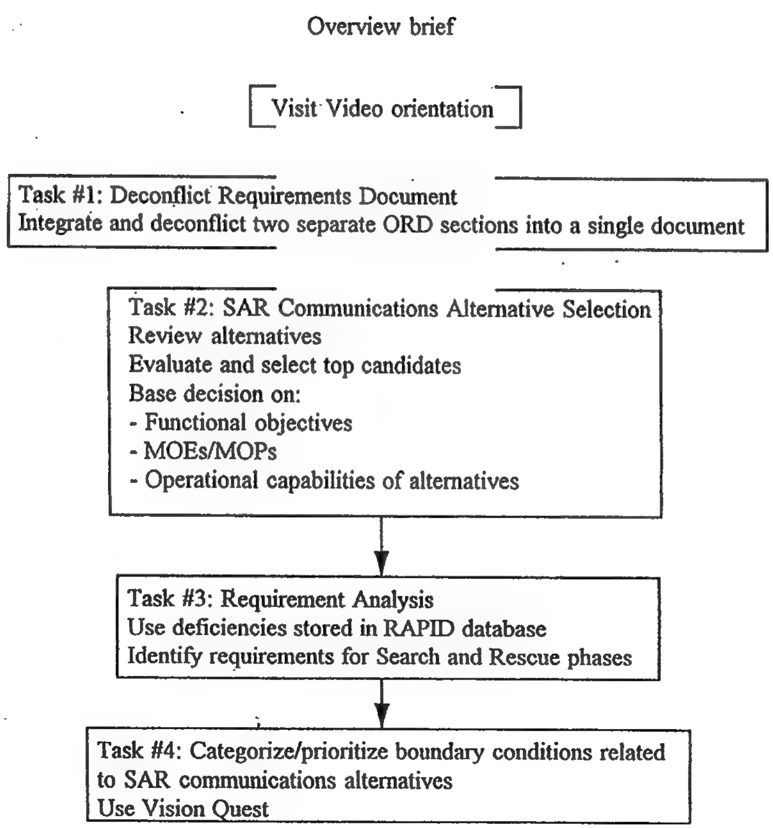


Fig. 5

Table 2

Pre-Milestone O-1 Collaborative Activities	Major Desktop Video Conferencing Activities	Time	Problems experienced, Help requested	Solutions found
Task #3: Search and Rescue Communications Alternatives Selection				
1. Group review and prioritization of alternatives	1.1 One site is creating the template for the alternative evaluation matrix in the screen share.	15 min	Editing table in screen share.	Solution for the shared matrix is illustrated in Fig.11
	1.2 The other site is analyzing the source document		Inability to save the copy back from screen share to the Word file.	
	1.3 Start searching the answers to the specified questions by looking at multiple resource documents.		Inability to look at the similar past cases	Search and rescue expert is guiding the search, while software person is typing the answers
	1.4 Type the answers to the evaluation table.			
	Activities 1.2 -1.4 continue till 12 alternatives are evaluated by 7 criteria	1:25		Results are retyped to the Word copy
	1.5 Selection of alternatives starts		They forget the reasons they used an hour ago to select the alternative (memory?).	
	1.6 Generation of comments on reasons to chosen selection starts			One site is generating the comments, the other is typing
	1.7 File transfer the upgraded Word copies	1:05		

Table 1

Utility Evaluation	**Generally experienced Action Officer view**	**Acquisition Expert View**	**Computer advanced Action Officer View**	**SAR Experienced Action Officer View**
1. The shared screen function provided by the VISIT software was understandable, with minimum training (interval scale)	3	2	4	3
2.The file transfer function aided in data acquiring for producing documents	3	4	4	4
3. The shared screen function enabled to more effectively communicate desired document changes		3	4	4
4. Overal VISIT seems to be a useful tool in the Requirements Generation Process	1&2	3	4	4
5. What improvements to VISIT would increase it's utility	See comment #1	See comment #2	See comment #3	See comment #4

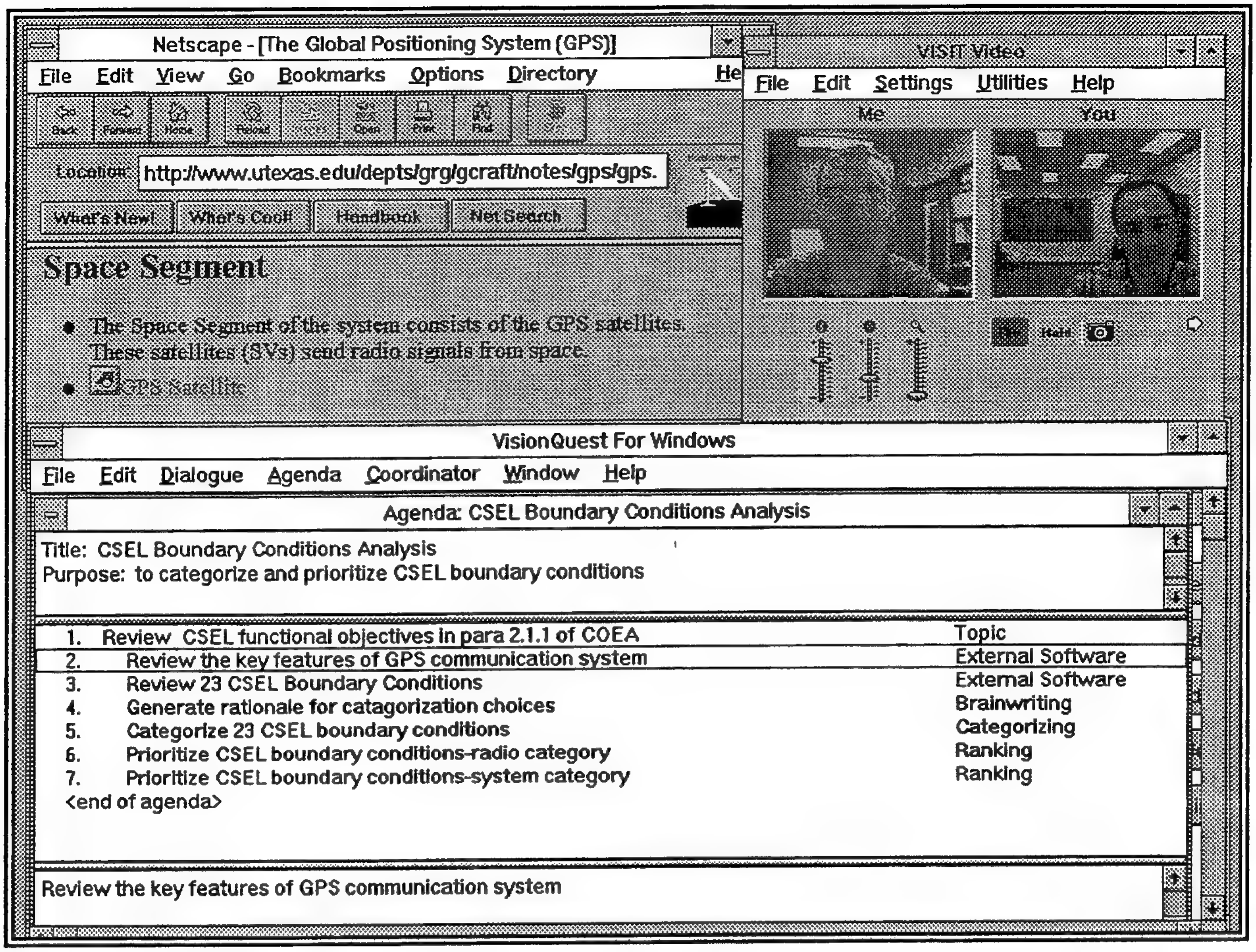

Netscape - [The Global Positioning System (GPS)]
File Edit View Go Bookmarks Options Directory He
Location: http://www.utexas.edu/depts/grg/gcraft/notes/gps/gps.
What's New! What's Cool! Handbook Net Search
Space Segment
The Space Segment of the system consists of the GPS satellites. These satellites (SVs) send radio signals from space.
GPS Satellite
VISIT Video
File Edit Settings Utilities Help
Me
You
VisionQuest For Windows
File Edit Dialogue Agenda Coordinator Window Help
Agenda: CSEL Boundary Conditions Analysis
Title: CSEL Boundary Conditions Analysis
Purpose: to categorize and prioritize CSEL boundary conditions
1. Review CSEL functional objectives in para 2.1.1 of COEA Topic
2. Review the key features of GPS communication system External Software
3. Review 23 CSEL Boundary Conditions External Software
4. Generate rationale for catagorization choices Brainwriting
5. Categorize 23 CSEL boundary conditions Categorizing
6. Prioritize CSEL boundary conditions-radio category Ranking
7. Prioritize CSEL boundary conditions-system category Ranking
<end of agenda>
Review the key features of GPS communication system

Fig. 6

Assessing Salience of Aspects of Real-World Images with a Neurophysiological Model of Early Vision

Marc L. Carter
Assistant Professor
Psychology Department

University of South Florida
Tampa, FL

Final Report for:
Summer Research Program
Armstrong Laboratory

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, Washington DC

and

Armstrong Laboratory

September 1995

Assessing Salience of Aspects of Real-World Images with a Neurophysiological Model of Early Vision

Marc L. Carter
Assistant Professor
Department of Psychology
University of South Florida

Abstract

Current models that predict the effects of laser exposure on human observers consider many factors, but one important factor that has been considered less often is the location of fixation relative to objects (and hence the laser source) in a scene, even though the location of fixation is itself a powerful determinant of laser exposure effects. The project described here develops a model for predicting the location of fixation in real-world scenes by combining several currently fashionable hypotheses about the structure and function of early vision in humans. The model functions in essentially three stages: The first stage employs spatially isotropic difference-of-Gaussian (DOG) filters to produce edge enhancement of objects in the scene. The second stage uses oriented Gabor filters at several spatial scales to produce maps of intensity-by-orientation; these maps are then combined into one map that shows blobs of varying size and intensity. A simple algorithm is then employed to rank-order the blobs in terms of intensity and area, thus providing a principled model of areas of high contrast within the scene, which areas are likely to be a determinant in the location of fixation.

Assessing Salience of Aspects of Real-World Images with

a Neurophysiological Model of Early Vision

Marc L. Carter

ILPEM (Integrated Laser Personnel Effects Model) is a computer model currently under development by the Optical Radiation Division of the Air Force Armstrong Lab (AL/OEO) for use in predicting the effects of laser exposure on human observers, especially in mission-critical tasks such as ground/air target identification, targeting of PGM, and aircraft visual identification, to name a few. ILPEM at present can simulate flashblindness, glare, and is currently being expanded to include laser-produced scotomata (both permanent and temporary), among others.

ILPEM functions by modeling parameters that mitigate or exacerbate laser bioeffects, and combining the component results to produce a sum effect; components modulating the effects of laser exposure include aircraft canopy, atmospheric and intraocular scatter, atmospheric attenuation, distance, type of laser, laser eye protection (LEP), and others. One factor that is also critical in the determination of laser effects on mission performance is the orientation of the eye with respect to the laser source (Green et al., 1988; Thomas, 1994). If the observer has the laser source at the center of gaze then the effects of the laser on mission performance will likely be much greater than if the source is eccentric to the axis of gaze, although the type of effect (small or large, temporary or permanent) will also be a consideration. Although direction of gaze is considered important, ILPEM contains no module that will predict the behavior of the observer's gaze in a given scene; it is the purpose of this project to develop a module that can be integrated

into ILPEM and that will predict the likely location of the pilot's gaze at the time of exposure.

In a sense this lacking component is one of visual search (either detection and/or identification), and there are several search models in the literature proposing to predict the location of a fixation (or of attention) with respect to a given visual display (see especially Howe, 1993; Nicoll, 1994). Most of these models, however, are highly constrained by the type of visual task (most have been developed using artificial displays), and the fit of these models to performance in realistic displays is less than adequate (Kosnik, 1995). In addition, most current models tend to use abstract mathematical representations for behavioral performance (e.g., Cook, Stanley & Hinton, 1995; Doll et al., 1995; Nicoll, 1994). In a model in which an aspect of a single domain is modelled (say search for contrast), using an abstract mathematical formalism to model performance is acceptable. However, modelling realistic observer performance under a variety of conditions (say, in conditions where chromatic contrast, spatial scale, or motion are relevant parameters of the search) will at some point necessitate the addition of other factors to the model, and as the model grows the mathematics and the subject of the model bear ever-decreasing resemblance to one another. It seems sensible, then, to begin a search model having ends in more general applications on a basis that is closely tied to the physiology and anatomy of the visual system, and using realistic displays (Biberman & George, 1994). In this way adding new features to the model can be accomplished via incorporation of new abilities to existing structure, rather than the addition of new mathematical structures. In addition, the integration of early and late visual processing components (as in models such as ILPEM) suggests a need for more realistic modeling of early visual processes, while at the same time preserving computational tractability (Garcia-Perez & Sierra-Vasquez, 1995).

Search can be divided into two abilities: detection and identification. Detection is a simple behavior, involving the registration of an event in the environment in the absence of any other more complicated processing; it is the simple realization that something is *there.* (One can make a case that detection in most circumstances requires discrimination, and I would agree with that, but it is a simpler process than identification.) Identification is the more complex task, including aspects of detection, discrimination, and categorization. It makes sense to begin the model with the goal of object detection, and later other abilities can be added. It should be noted that detection can be accomplished along any of several visual dimensions, including simple contrast, orientation, spatial scale, shape, color, and motion, as well as relationships between these dimensions, to name a few.

The general thrust of any model of detection in visual search will depend almost entirely on some measure of salience or conspicuity of the objects in view. After all, what determines where we look or attend is largely what is there, and what is "there" is a function of both the sensitivity of the visual system to objects at visual locations, as well as the distribution of objects across visual locations. One measure of conspicuity that has received much recent interest is spatial contrast, and there are several models which can assess the contrast of objects in a scene and thus lead to a relatively objective measure of salience (Cannon, 1995; Peli, 1995), and there is good evidence that the human spatial contrast sensitivity function predicts detectability of real-world images (Owsley & Sloane, 1987; Thomas, Brakefield, & Barsalou, 1992). The model described here is in this tradition, but deviates from those earlier models in that it relies heavily on the anatomy and physiology of the visual system in developing a measure of contrast, one that initially includes spatial contrast at many spatial scales and orientations (in the vein of Campbell,

Graham, & Wilson; see Campbell & Robson, 1968; Graham & Nachmias, 1971; Wilson, 1991; Thomas, Brakefield, & Barsalou, 1992). The model requires no high-level symbolic representations; neither does it rely on intensive training as would be necessary in a connectionist net designed to perform object (target) recognition (e.g., Doll et al., 1995). Its purpose is to provide a measure of a realistic scene's "visually interesting" aspects, and thus provide information about where in such a scene an observer would tend to look, by substituting mathematical approximations of the physiological and anatomical components of the visual system likely to be important in visual search.

The model at its current stage of development has three parts. The first part is strictly an image-processing system, one that extracts and processes information from a gray-scale intensity map (a black-and-white image), in two stages. From this the model produces blobs of varying intensity, which are fed into and algorithm designed to provide estimates of the location of a fixation with respect to the weighted sum of these oriented contrast measures. From these probabilities we can predict where a fixation is likely to be (in serial order), and with a few assumptions about the length of fixations, we can predict a most likely location for fixation at given points in time after the scene is presented to the observer.

The following report will present the motivation and logic of the model, and then go on to discuss the mathematical approximations used in the simulation. Following this will be a short description of the current state of the model (as of 4 August 1995), and finally an outline for the continued development and completion of the model.

Visual search is a process that has received a great deal of study in the past few years.

Search has been studied as a process, and from time to time as a diagnostic for the study of low-level visual information processing. It seems to be a consensus among researchers that visual search is a low-level phenomenon, one that takes place in the first few stages of the processing of visual information, and almost surely relies heavily on the structure and processes in primary visual cortex (V1), and perhaps the next few stages.

The earliest stages of vision have often been modeled as a series of filters, analyzing the visual information into various components, including color, motion, shape, size, and others. The analysis stage of vision would then likely (but not necessarily) be followed by a synthetic stage, one which would represent the aggregation of the earlier, analytic stage.

The model begins with input as a gray-scale (intensity) map, since only 5-10% of the contrast measurable at the level of photoreceptors is chromatic contrast (Geisler, 1995). The input arrays are 256 x 512 pixel values, arrayed rectangularly, with uniform density of samples; i.e., there are no model parameters for eccentricity, either for the change in photoreceptor type and density or cortical magnification; since the model is initially designed to produce a probable location for the center of gaze, later components can be implemented to produce effects of eccentricity. Currently the input to the model is an abstract representation of intensity changes (geometrical shapes), but the model is designed to operate on real-world images.

The model has three chief elements in this first stage. First is a function to mimic the optical spread function of the human eye, following Geisler (1995):

$$(\,.952 * \exp\{-2.59 * |\rho|^{1.36}\}) + (\,.048 * \exp\{-2.43 * |\rho|^{1.74}\})$$

Although largely manifest as a defect of the optics of the eye, the spread of a point source at the retina serves to effectively increase the resolution of the eye without producing aliasing. The function in the model serves the same purpose, roughly low-passing the input image to remove high-spatial-frequency components that would produce aliasing in the outputs of the high-spatial-frequency filters.

The next stage of processing is designed to implement retinal processing. The dominant structure of the receptive fields of retinal ganglion cells is the of a center-surround antagonism, with both center-on and center-off cell types. Among the cells of the P (X) system, there are cells with sensitivity for chromatic contrast, and among the M (Y) cells there is a dominance for luminance contrast. At present the model considers only luminance contrast, and implements these receptive fields with difference-of-Gaussian (DOG) functions at six spatial scales to provide a sharpening of the edges of the image that have been blurred by the optical spread function. This stage also serves to help produce a reduced sampling rate for the low-frequency filters in the next stage. The functional form of the DOG filters in the model is:

$$D(x,y) = c\left(\mathbf{k}_1 e^{-\left(\frac{x^2+y^2}{2\sigma_1^2}\right)}\right) - \left(\mathbf{k}_2 e^{-\left(\frac{x^2+y^2}{2\sigma_2^2}\right)}\right)$$

where σ_1 and σ_2 are the variances of the center and surround sections, respectively, and k_1 and k_2 are weighting factors. Following Marr (1982), the ratio of surround to center σ's is approximately 1.6, and at present the *k* values are set to produce unit Gaussians. The factor *c* is used to scale center/surround so that the center section also produces a DC component in the output that represents pure intensity or activity in the receptors at that location. The result of this stage is a set of six filtered images, low-passed and differentiated to produce enhancement of

the contrast at edges and thus highlight intensity difference in the input, and serve as input to the next stage. This bank of filters also removes very-low-spatial frequency changes that would occur with changes in the illuminant rather than changes in reflectance due to objects in the scene, and also helps serve as a gain control.

The last component of the retinal processing is a half-wave rectification of the DOG signals, and exponentiation of the response. At present the model simply uses an RMS value (square the output and exponentiate by 0.5) to produce a rectified signal, which is then passed to the next stage.

The next stage represents the processing that occurs early in cortex, probably at the level of V1. Again, there are three components to this section of the model, aimed at representing information that would be most likely used in attracting an observer's gaze.

Several models of the response of cortical simple cells have been proposed, and usually converge on a Gaussian-tapered sine or cosine function. This is most commonly modeled as a gabor or wavelet function, oriented and with varying spatial frequencies. In the model this is implemented by a bank of oriented cosine Gabor filters to simulate the recovery of oriented edges and objects at different spatial scales. The form of the filters is

$$F_{\omega,\theta}(x,y) = k e^{-\left(\frac{x^2+y^2}{2\sigma^2}\right)} \cos(2\pi\theta\omega x)$$

with ω and θ the spatial frequency and orientation of the filter, respectively. The model has six spatial frequencies, from 0.5 to 16.0 cycles/image width, each center frequency differs from the adjacent frequency by an octave; there are eight orientations, from 0 to 157 degrees, equally

spaced. The variance of the filters is one-third the center frequency, and k is again a scalar used to set the area under the Gaussian to unity.

The gabor filter bank has a number of features that are based on the known (and/or assumed) physiology: they provide lateral inhibition in the spatial domain, and exhibit a behavior qualitatively similar to V1 complex cells. A second stage of processing takes place between the oriented filters, or inhibition in the spatial frequency domain (DeValois & DeValois, 1990), behavior that has been shown to occur by a release from lateral inhibition effects (DeValois, 1977; Tolhurst & Barfield, 1978). This is accomplished in the model by the following equation:

$$\mathrm{F}_{\theta_i \omega_j}(x,y) = \mathrm{F}_{\theta_i \omega_j}(x,y) - \alpha\left(\mathrm{F}_{\theta_i \omega_{j-1}}(x,y) + \mathrm{F}_{\theta_i \omega_{j+1}}(x,y)\right)$$

where (again) θ and ω are orientation and spatial frequency, and α is a weighting factor.

The last component of the cortical section is a gain function applied to the outputs of the filters. Since the human contrast sensitivity function shows that humans are not uniformly sensitive to all spatial frequencies, the outputs in the filter maps are weighted by the following function:

$$\mathrm{F}_{\theta,\omega}(x,y) = 168\, \nu\, e^{(-2\nu)}$$

where ν is the spatial frequency of the filter in cycles per degree. This forces the peaks to conform roughly to the form of the human CSF. At present, since the model uses rectangular arrays, there is also a weighting factor for orientation designed to render equivalent the sensitivity of the filters at all orientations (ignoring, and in fact removing, the oblique effect--an oblique effect is a natural byproduct of rectangular arrays). Another function for the gain controls would be to simulate attentional sets: We could, for instance, bias the model to be more sensitive to

vertical and horizontal orientations, or certain spatial scales, to simulate the effects of an observer attentively "looking" for regular objects of certain size in an otherwise unstructured display. Thus more energy would be output in the filters of those orientations and scales, and thus provide a crude representation of search for objects, without assuming any more intelligence on the part of the system than a bias for certain low-level attributes of the scene.

A last step in the model is a sum of the outputs of the filters to produce a map that yields total activity: intensity, contrast, orientation, spatial scale are all represented at each location in a map. An algorithm is then applied to the map; this algorithm probabilistically, but with weights set by the activity levels of the regions in the display, selects and rank orders the bright regions in the display; the algorithm integrates area by intensity (similar to Bloch's Law) and these are weighted by distance of the nearest point to the center of the display (assuming for the present that fixation begins in the center of the display). This is then a representation of the areas of the display that would be most likely, and in what order would be selected, for fixation by a human observer.

Further Work.

The most important next step involves the use of real-world imagery and empirical validation of the model response. compared to human observers. Ideally the human observers will report more quickly or more frequently the objects that the program indicates to be the most salient. If not, then there are several ways in which the model can be tuned to more closely mimic the performance of human observers. After this is accomplished we can then move toward

integrating the model with ILPEM.

Other additions will include chrominance (color) contrast, which can be gotten via the use of the colors in an image simply by extracting RGB values and feeding them into an opponent process preprocessor. This has already been done successfully in other models (Cook et al., 1995; Doll et al., 1995). Then we would follow the processing of the program as it is now, with the addition of color-contrast maps. Motion contrast sensitivity can be implemented by the same algorithm that supplies the gabor filtering, by applying gabors across the time rather than the spatial domain. The addition of motion will also require the implementation (or integration) of temporal processes, such as adaptation (see, e.g., Hahn & Geisler, 1995). There are already modules in ILPEM that contain values for the assessment of effects as they change over time; these might best be used in the present model, once motion is included.

Currently the model is rather incomplete, but is very nearly ready to integrate into ILPEM. I would suppose that after the model is built into ILPEM the other additions and extensions can then be added. This is not, and perforce cannot be, a complete model of early visual processing. But it can serve as a tool to help researchers understand the most likely location of an observer's fixation point in a real-world scene, and this will enable us to more accurately assess the effects of laser exposure on mission performance.

References

Biberman L.M., & George, V. (1994). The use of target dimensions in predicting probability of recognition. *Proc. IRIS Passive Sensors, II,* 213-224.

Campbell, F.W. & Robson, J.G. (1968). Application of Fourier analysis to the visibility of gratings. *Journal of Physiology (London), 197,* 551-566.

Cannon, M. W. (1995). A multiple spatial filter model for suprathreshold contrast perception. In E. Peli (Ed.), *Vision Models for Target Detection and Recognition,* pp. 88-116. Singapore: World Scientific Publishing.

Cook, K.J., Stanley, P.A., & Hinton, J.L. (1995). The Oracle approach to target acquisition and search modelling. In E. Peli (Ed.), *Vision Models for Target Detection and Recognition,* pp. 135-171. Singapore: World Scientific Publishing.

DeValois, K.K. (1977). Spatial frequency adaptation can enhance contrast sensitivity. *Vision Research, 17,* 1057-1065.

DeValois, R.L. & DeValois, K.K. (1990). *Spatial Vision.* Oxford: Oxford Science Publications.

Doll, T. J., McWhorter, S. W., Schmeider, D. E., & Wasilewski, A. A. (1995). Simulation of selective attention and training effects in visual search and detection. In E. Peli (Ed.), *Vision Models for Target Detection and Recognition,* pp. 396-418. Singapore: World Scientific Publishing.

Garcia-Perez, M. A. & Sierra-Vasquez, V. (1995). Visual processing in the joint spatial/spatial-frequency domain. In E. Peli (Ed.), *Vision Models for Target Detection and Recognition,* pp. 16-62. Singapore: World Scientific Publishing.

Geisler, W. S. (1995). Discrimination information in natural radiance spectra. In E. Peli (Ed.),

Vision Models for Target Detection and Recognition, pp. 117-131. Singapore: World Scientific Publishing.

Graham, N. & Nachmias, J. (1971). Detection of grating patterns containing two spatial frequencies: A comparison of single channel and multichannel models. *Vision Research, 18,* 815-825.

Green, R. P., Cartledge, R. M., Cheney, F. E., & Menendez, A. R. (1990). Medical management of combat laser eye injuries. *USAFSAM-TR-88-21R.*

Hahn, L. W., & Geisler, W.S. (1995). Adaptation mechanisms in spatial vision--I. Bleaches and Backgrounds. *Vision Research, 35,* 1585-1594.

Howe, J.D. (1993). Electro-optical imaging system performance prediction. In M.C. Dudzik (Ed.), *Electro-optical Systems Design, Analysis and Testing,* pp. 55-120.

Kosnik, B. (1995). Quantifying target contrast in target acquisition research. In E. Peli (Ed.), *Vision Models for Target Detection and Recognition,* pp. 380-395. Singapore: World Scientific Publishing.

Marr, D. (1982). *Vision.* New York: W.H. Freeman & Co.

Nicoll, (1994). A mathematical framework for search: The neo-classical model. *Proc. IRIS Passive Sensors, II,* 171-189.

Owsley, C. & Sloane, M. E. (1987). Contrast sensitivity, acuity, and the perception of 'real-world' targets. *British Journal of Ophthalmology, 71,* 791-796.

Peli, E. (1995). Simulating normal and low vision. In E. Peli (Ed.), *Vision Models for Target Detection and Recognition,* pp. 63-87. Singapore: World Scientific Publishing.

Thomas, S. R. (1994). Aircrew laser eye protection: Visual consequences and mission

Vision Models for Target Detection and Recognition, pp. 117-131. Singapore: World Scientific Publishing.

Graham, N. & Nachmias, J. (1971). Detection of grating patterns containing two spatial frequencies: A comparison of single channel and multichannel models. *Vision Research, 18,* 815-825.

Green, R. P., Cartledge, R. M., Cheney, F. E., & Menendez, A. R. (1990). Medical management of combat laser eye injuries. *USAFSAM-TR-88-21R.*

Hahn, L. W., & Geisler, W.S. (1995). Adaptation mechanisms in spatial vision--I. Bleaches and Backgrounds. *Vision Research, 35,* 1585-1594.

Howe, J.D. (1993). Electro-optical imaging system performance prediction. In M.C. Dudzik (Ed.), *Electro-optical Systems Design, Analysis and Testing,* pp. 55-120.

Kosnik, B. (1995). Quantifying target contrast in target acquisition research. In E. Peli (Ed.), *Vision Models for Target Detection and Recognition*, pp. 380-395. Singapore: World Scientific Publishing.

Marr, D. (1982). *Vision.* New York: W.H. Freeman & Co.

Nicoll, (1994). A mathematical framework for search: The neo-classical model. *Proc. IRIS Passive Sensors, II,* 171-189.

Owsley, C. & Sloane, M. E. (1987). Contrast sensitivity, acuity, and the perception of 'real-world' targets. *British Journal of Ophthalmology, 71,* 791-796.

Peli, E. (1995). Simulating normal and low vision. In E. Peli (Ed.), *Vision Models for Target Detection and Recognition*, pp. 63-87. Singapore: World Scientific Publishing.

Thomas, S. R. (1994). Aircrew laser eye protection: Visual consequences and mission

Employing Analogies to Enhance Conceptual Knowledge Acquisition

Richard Catrambone
Associate Professor
School of Psychology

Georgia Institute of Technology
Atlanta, GA 30332

Final Report for:
Summer Faculty Research Program
Armstrong Laboratory

Sponsored by:
Air Force Office of Scientific Research
Brooks Air Force Base, San Antonio, Texas

and

Armstrong Laboratory

September 1995

Employing Analogies to Enhance Conceptual Knowledge Acquisition

Richard Catrambone
Associate Professor
School of Psychology
Georgia Institute of Technology

Abstract

An approach to aiding the acquisition of conceptual knowledge has been developed that is being applied to the Stat Lady probability and statistics computerized training environment. This approach uses analogies to help learners understand concepts. A careful decomposition of the curricular elements taught via Stat Lady produced a breakdown of the to-be-learned knowledge into three types: symbolic, procedural, and conceptual. Prior work has shown that the Stat Lady environment is useful for aiding the acquisition of symbolic and procedural knowledge, but not as effective in aiding the acquisition of conceptual knowledge. Relevant research has suggested that learners can acquire new conceptual knowledge more effectively if the new knowledge is systematically tied to pre-existing knowledge through the use of analogies. Thus, the work described in this document tests a principled approach to aiding the acquisition of conceptual knowledge in a computerized training environment.

Employing Analogies to Enhance Conceptual Knowledge Acquisition

Richard Catrambone

Introduction

You have to learn to walk before you can run. Or do you? This simple metaphor for acquiring new knowledge and skills has been implemented in countless pedagogies, but may not be completely true. For instance, in ACT-R (Anderson, 1993) all knowledge starts in declarative form and productions (for carrying out procedures) are formed through an analogy process. Similarly, in SMART (Shute, in press), the underlying learning theory posits a uni-directional link from symbolic knowledge (SK) acquisition (i.e., basic, abstract curriculum elements) to procedural skill (PS) acquisition (i.e., the application of that knowledge).

Two learning questions arise though. First, how immutable is this "uni-directional" connection between SK and PS? For instance, how would learning be affected by presenting the learner with conceptual knowledge (CK) prior to SK or PS? Current findings in the literature suggest that conceptual knowledge, presented at the *outset* of instruction, may actually enhance ensuing understanding and problem solving activities (Chi, 1993). SMART asserts that increased conceptual knowledge (CK) can feedback to influence procedural skill (PS) acquisition. Thus, one of the goals is to examine whether CK instruction can aid acquisition of SK and/or PS.

A second learning issue concerns ways to facilitate the acquisition of conceptual knowledge. There is growing support in the literature for the notion that analogies, if created and employed properly, can be used to explicitly instruct CK, and consequently influence both CK and PS acquisition (e.g., Gentner & Gentner, 1983; Kieras & Bovair, 1984). A second goal then is to examine the degree to which we can generate and effectively utilize analogies, within the domain of descriptive statistics, to enhance CK acquisition.

A third, and related, issue is to explore the theoretical basis for examining the efficacy of single versus multiple analogies in conveying CK. That is, is a concept (e.g., central tendency) learned better when a single, unified analogy is used to illustrate its various facets (e.g., mean, median, and mode) or when separate, multiple analogies are used to illustrate those facets? Perhaps fine-tuned analogies for each facet are better than a single analogy that does an adequate job on each facet, but not as good as each of the fine-tuned ones. This analysis also depends on whether, in a particular situation, the analogy is worth the trouble. Presumably an analogy has value if the base analog or domain (e.g. a seesaw) is better known by the learner than the target analog (e.g., the mean). If the base analog is relatively easily understood, then the teacher must be concerned with making sure the mapping is relatively straightforward.

Empirical Motivation

Shute (in press) conducted two experiments testing the contribution of SMART (Student Modeling Approach for Responsive Tutoring) to learning outcome and efficiency. SMART tracks whether a learner has acquired a particular curricular element and provides error feedback that is appropriate to the learner's presumed state of knowledge. In Study 1 (Shute, in press), SMART was disabled, and in Study 2, the program was enabled. Two significant 3-way interactions were found involving learning gain × outcome type × aptitude level, both having direct relevance to the current project (see Figure 1). In general, low-ability learners showed dramatic pretest-posttest improvements for SK and PS, but only moderate gains for CK; high-ability learners improved equally across all outcome types.

This finding motivated the current research to find some means of similarly boosting CK acquisition, especially for low-ability learners. To achieve this goal, an experiment has been designed to: (a) test the effects on learning when the different outcome types (SK, PS, and CK) are instructed in different orders, (b) examine the effects of explicitly instructing conceptual knowledge (CK) via analogies, and (c) investigate the effects of employing a unified analogy versus multiple

analogies for instructing measures of central tendency. We believe that analog instruction may ultimately enhance the memorability of CK elements and also help learners acquire related PS elements.

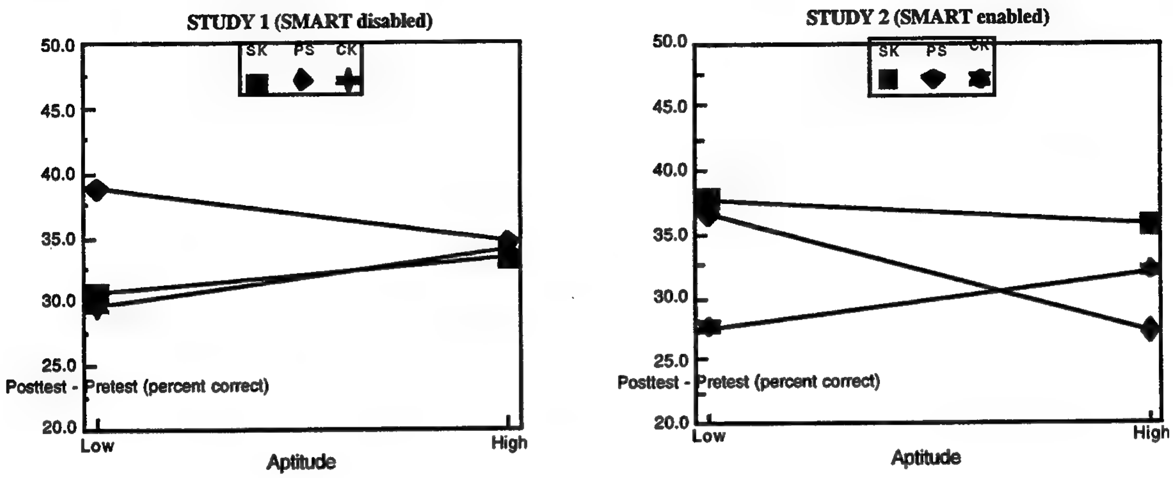


Figure 1. Three-Way Interactions: Learning Gain × Outcome Type × Aptitude Level

<u>Project Description</u>

Stat Lady--Descriptive Statistics (DS)--consists of three modules. The first module (DS-1) instructs data organization and plotting, the second (DS-2) teaches issues of central tendency, and the third (DS-3) teaches variability. DS-1 is currently completed and has been used in a series of controlled experiments (Shute, Gawlick-Grendell, Young, & Burnham, in press). DS-2 is the module being used in the current project, and DS-3 is still in progress. All three modules follow the same basic structure in that curricular elements (CEs), for SK, PS, and CK outcome types, are extracted from a cognitive task analysis of the domain, and are then instructed and assessed within Stat Lady.

<u>Pedagogy</u>. To illustrate, Stat Lady instructs a particular CE (or small group of related CEs), such as the PS element representing the ability to compute the mean. Subsequently, learners are presented a problem set designed to assess their understanding. Learners obtain data from the "Number Factory" and then proceed, step-by-step, to solve the problem that requires them to

compute the mean from that unique set of data. Help (in the form of three-level feedback) is provided by Stat Lady as needed. The system updates its record about whether learners have learned particular CEs.

CE-Derivation. We began this project by systematically decomposing the descriptive statistics curriculum into a hierarchically-related set of CEs that were subsequently classified into three outcome types. We then created analogies for instructing the higher-level CK elements (in this case, mean, median, and mode). For example, following more formal (i.e., lower-level, symbolic knowledge) instruction on the mean, an on-line seesaw appears with numbers arrayed along the board, similar to the x-axis of a graph with differing heights representing different frequencies. Moreover, the seesaw consists of a moveable fulcrum so that learners can slide the fulcrum, horizontally under the board, until the board (tipping appropriately in relation to the fulcrum) becomes balanced (see Figure 2). While the fulcrum is moving, numbers dynamically appear within the fulcrum representing the corresponding point/value along the board (x-axis). The number that shows up when a balance state is attained reflects the mean for those particular data.

CK Instruction via "Unified" Analogy. The mapping between the seesaw and the mean would be as follows: To "balance" a seesaw (base domain), one needs to distribute the "weights" such that they occur "equally" on either side of the fulcrum. To achieve this state, either the weights can be moved around, or the fulcrum. The weights represent different (magnitudes of) numbers and the mean (target domain) maps onto the fulcrum. Learners must be given the opportunity to actually play with the seesaw by moving the fulcrum.

The mapping between the seesaw and the median is less straightforward. The seesaw would become balanced (i.e., horizontal) when the fulcrum is moved to the median value. In Figure 2 this value would be 6. It may look somewhat strange to the learner to have the seesaw become balanced at such a spot.

Finally, the mapping between the seesaw and mode is even less clear. In this case the seesaw will balance when the fulcrum is under the tallest pile of boxes. This situation could look very odd and violate the expectations of most learners.

However, the value of using such a unified analogy for mean, median, and mode may outweigh this potential disadvantage. The mismatch between the physical reality (balancing only at the mean) and the demonstration (balancing at median and mode) is precisely the point. That is, when the learner sees the seesaw balancing at the mean and then later balancing not at the mean, but rather at the median or mode, may make the distinction among the concepts that much clearer.

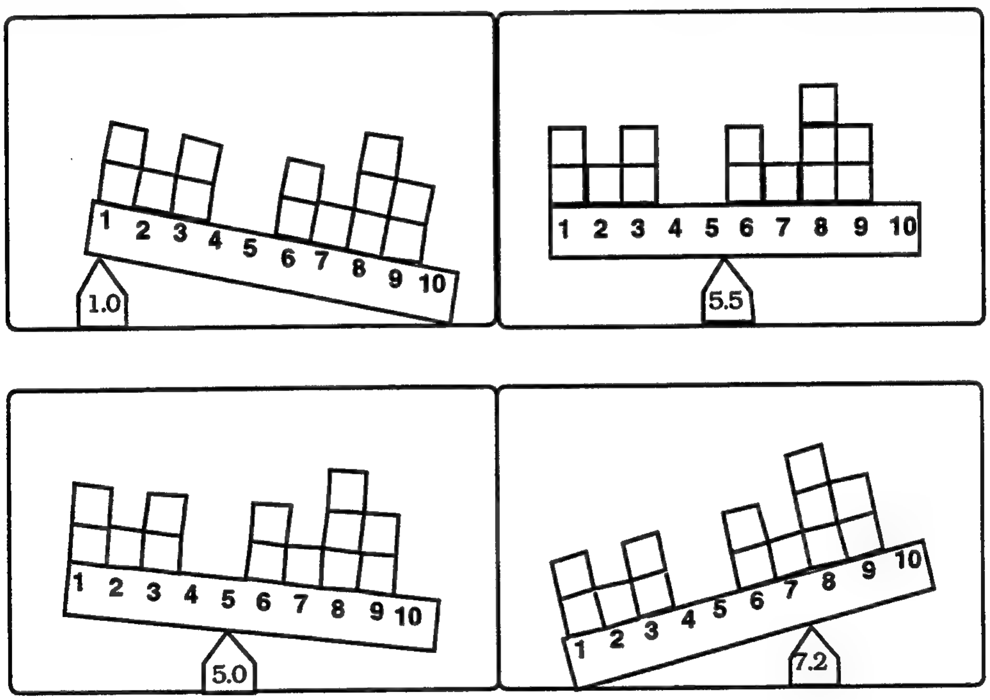


Figure 2. A Seesaw Analogy Illustrating the Mean

<u>CK Instruction via Multiple Analogies</u>. In contrast to the above argument for the value of a unified analogy to teach related concepts, perhaps analogies that are specially tailored to each facet of a concept would be more useful. Each analog could provide a mapping to the relevant facet that might be more straightforward than the mapping involved in using a unified analogy.

For instance, Figure 3a represents a "folding number line" that provides a clear mapping between position on the line and the median. Figure 3b contains "hanging buckets of balls" that shows a mapping between frequency (number of balls) and how far a bucket stretches.

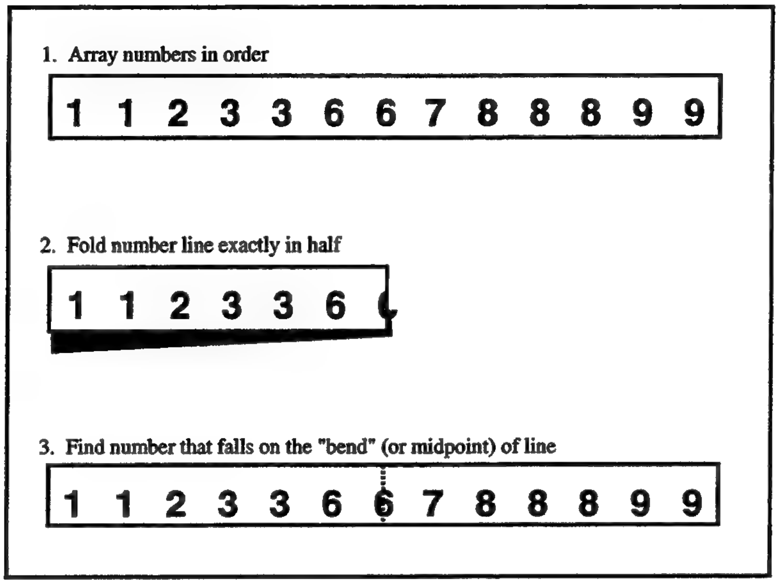


Figure 3a. Folding the Number Line, Illustrating the Median

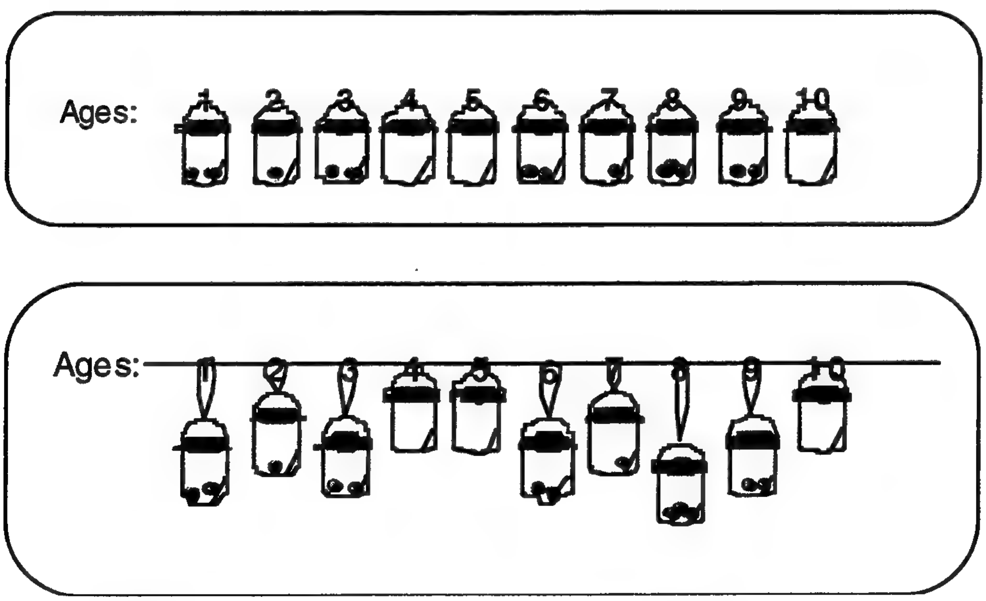


Figure 3b. Hanging Buckets of Balls, Illustrating the Mode

In summary, there are three main research questions that we are examining: (a) Order effects related to outcome types--Is it better to begin instruction on a topic with conceptual knowledge to establish a foundation, followed by more formal knowledge (SK), or vice versa? (b) Analogy effects on CK acquisition--Is conceptual knowledge acquisition facilitated as a function of using analogies for instruction/remediation in contrast with a condition involving no analogies, and does this ultimately enhance SK and PS acquisition as well? (c) Effect of unified versus specially-tailored analogies on CK learning gains--Is transfer enhanced (or thwarted) by employing a single (unified) analogy vs. multiple (tailored) analogies to instruct the topics of interest?

Experiment

For this study, we chose to instruct only a circumscribed portion of the original Stat Lady DS-2 curriculum (i.e., mean, median with N = odd, and mode). Other CEs that were originally targeted for instruction will be used instead as measures of transfer (e.g., skewness, median with N = even). That is, in DS-2, Stat Lady initially instructed and assessed the median for both odd and even numbers (the latter representing the more complex of the two). But to assess learning gains as well as transfer, we need to remove certain CEs from the curriculum (explicit instruction) for inclusion on the posttest to gauge transfer.

Figure 4 depicts the 12 basic conditions (and sample sizes per cell) in the upcoming study (between-subject design). As mentioned, the curriculum will be ordered: mean, median (odd) and mode, and within each, the CEs (corresponding to the different outcome types) will be arranged as shown below. SK and PS elements will remain instructed and assessed as they were in the original version of the Stat Lady DS-2 curriculum (except when holes need to be filled), and CK elements will be instructed (a) as they were, or (b) as they were plus analogies added to instruction. There are no conditions in which PS is instructed first because, in line with current learning theory, knowledge and skill acquisition start with declarative knowledge (i.e., either SK or CK; Anderson, 1993).

The three treatment conditions for this study are thus: (a) *Instruction Only*--presentation of the instructional material as is, without any analogies presented (the "None" column in Figure 4); (b) *Unified Analogy*--instruction of the three main concepts as in the Instruction Only condition, but additionally using a single analogy (i.e., a seesaw with manipulable fulcrum) to illustrate the concepts, presented subsequent to the standard CK instruction; and (c) *Multiple Analogies*--instruction of the three topics like the Instruction Only condition, but using three different analogies to illustrate the main concepts: Mean = Seesaw (like in the Unified Analogy condition), Median = Folding Number Line, and Mode = Hanging Buckets of Balls. Again, all analogies will be presented following the standard CK instruction.

			Analogies in CK Instruction		
Order by Outcome Type			None	Unified	Multiple
SK	PS	CK	N = 40	N = 40	N = 40
SK	CK	PS	N = 40	N = 40	N = 40
CK	SK	PS	N = 40	N = 40	N = 40
CK	PS	SK	N = 40	N = 40	N = 40

Figure 4. Treatment Conditions for the Analogies Study (N = 480)

Equating Time Spent on Task

Participants in the analogy conditions will be receiving additional instruction and spending more time on task compared to learners who will not receive the analogy instruction. In order to control for this, no-analogy participants need to receive extra practice on those components of the curriculum on which the analogy participants will receive the analogies. This will be accomplished by providing the no-analogy participants with additional exercises that deal specifically with the

CK components that are presumably being enhanced by the analogies. Pilot testing will be done so that the time on task measure is roughly equated for these two conditions.

The additional practice given to the no-analogy participants is also partly motivated by the assumption often made by educators that more practice in problem solving yields positive returns in terms of transfer to novel problems (although Sweller (e.g., Sweller & Cooper, 1985) has found little empirical support for this claim).

Transfer Tasks

The transfer tasks administered to participants will test their ability to extend what they have learned to novel cases. Each class of novel problems is described below:

(a) *Find median when n is even.* This will involve the learner being able to infer that one must take the mean of two values when there is no single middle value. This problem will require that the two middle values not be the same number.

(b) *Find modes in a bimodal distribution.* This will involve the learner recognizing that a distribution can have more than one mode. We will also ask learners to identify the mean and median from a graph of a bimodal distribution to see if there is any confusion/generalization concerning the fact that if there can be two modes, then perhaps there can be two means or medians.

(c) *Test whether the notions of mean, median, and mode are meaningful for different types of data* (e.g., categorical data).

The following problems test transfer in that, while similar to problems seen during instruction, they do not provide the scaffolding given during instruction. That is, the learner is not guided to use the "correct" approach, but rather, is now being tested to see if he or she will use it when applicable.

(d) *Find median and see if counting or formula is used.* This occurs in the context of large and small data sets (appearing as ordered lists).

(e) *Find median and observe if counting or cumulative frequency is used.* Provide the learner with a frequency table that includes a column for cumulative frequency and observe if the leaner counts to find the median, or takes advantage of the cumulative frequency column.

(f) *Determine method of computing the mean.* Given data sets involving repeated values or no repeated values, see which approach the learner uses to calculate the mean.

(g) *Test knowledge of the mean's relationship to its underlying distribution.* Given a problem involving two samples with equivalent means, the learner is asked to indicate which sample is more likely to contain cases equaling or exceeding a certain value. This gets at issues related to skewness.

The following problems involve more conceptual leaps and perhaps might be especially apt at testing benefits (and pitfalls) of analogies (as well as unified versus multiple analogies).

(h) *Determine selection of appropriate measure of central tendency in different contexts.* Given problems showing different distributions, the learner must decide whether the mean, median, or mode is the most representative measure of central tendency. For instance:

> Suppose you are trying to decide between two equally desirable jobs. The only issue you are concerned about is salary. For job A, the mean salary in that company is $100,000/year, and in job B, the mean salary is $80,000/year. Pick the pair of distributions below that would lead you to take job B over job A (even though the mean salary for job A is greater than the mean salary for job B). [The presentation of this problem will require displaying a few distribution pairs including one in which the median for the company with job B is greater than the median for the company with job A]

(i) *Determine inferences about flat distributions.* Given a picture of a flat distribution, the learner is asked what it has in common with the normal distribution with respect to mean, median, and mode. The correct answer is that, in both cases, the three measures will equal each other.

Method

Description of 12 Experimental Conditions[1]

(1) **SK-PS-CK$_{NA}$.** Within each of the three measures of central tendency (mean, median, mode), the ordering of outcome types for instruction are: SK, PS, CK. No analogies are used to instruct CK. To illustrate, SK related to the mean is instructed by direct presentation of relevant symbols, rules, and formulas; PS is instructed via demonstrations and interactive instruction; and CK is instructed via discussion of the relationships underlying the particular concept. Learners additionally are expected to induce functional relationships themselves. This represents the current default condition--Stat Lady's typical pedagogical approach.

(2) **SK-PS-CK$_{UA}$.** This condition is the same as the previous default condition except that Stat Lady instruction employs a single analogy (the seesaw) to teach the CK elements associated with the three measures of central tendency.

(3) **SK-PS-CK$_{MA}$.** This condition represents the same ordering as the other two conditions, only instead of using a single analogy (seesaw) to instruct the measures of central tendency, Stat Lady presents alternative analogies, per concept. These "multiple analogies" have been specially tailored to fit each concept.

(4) **SK-CK$_{NA}$-PS**. This condition, as well as the following two, represent a twist on the default pedagogical approach (i.e., conditions 1-3). Specifically, formal knowledge (SK) is first presented, followed by the associated CK (without analogies). Both of these precede PS instruction, and are presumed to lay the foundation for (and possibly enhance) PS acquisition.

(5) **SK-CK$_{UA}$-PS**. This represents the same ordering as condition 4, with a unified analogy being added to the curriculum to teach CK elements.

(6) **SK-CK$_{MA}$-PS**. This the same ordering as conditions 4 and 5, but it differs in that multiple analogies are employed to teach the CK.

(7) **CK$_{NA}$-SK-PS**. In this, and the next two conditions, CK elements are instructed first, followed by more formal instruction of the rules and/or formulas (SK), and ending with instruction

and practice on the PS elements. Thus, a conceptual foundation is established prior to the more formal formulas/rules. This is similar to an approach taken by Shute & Glaser (1991) in Smithtown where CK was the primary focus, and students had to learn microeconomic principles from an exploratory microworld. After each concept had been acquired, the formal rule (SK) was presented (e.g., "Congratulations. You've just discovered what economists refer to as the law of demand").

(8) **CK_{UA}-SK-PS**. This condition is the same as the preceding one, only CK instruction is supplemented by using a unified analogy across the three main concepts.

(9) **CK_{MA}-SK-PS**. This condition is the same as the preceding one, only CK elements are instructed by employing multiple analogies per concept.

(10) **CK_{NA}-PS-SK**. This condition, and the following two, comprise an approach that differs from the three preceding ones (1-3, 4-6, and 7-9) in that CK is taught initially, followed by PS instruction. This provides learners with an opportunity to immediately apply the developing CK. Subsequently, learners are given the relevant, formal rules/formulas (SK) at the end of the instructional segment.

(11) **CK_{UA}-PS-SK**. This condition is the same as above, only the initial CK is additionally instructed with unified analogies.

(12) **CK_{MA}-PS-SK**. This condition is the same as above, only the initial CK is instructed using multiple analogies.

Hypotheses and Experimental Tests

Order Effects. We hypothesize that condition 7 (or the cluster of conditions 7, 8, and 9) will show the highest outcome scores overall. Specifically, we are positing that it is more efficacious to initially provide a conceptual foundation, then present the formal symbolic knowledge, followed by the provision of practice opportunities. Not only should CK be enhanced, but also PS, both in terms of learning outcome and efficiency. The specific conditions to be compared are: 1, 4, 7, and 10 (i.e., the four main order conditions without analogies presented).

More particular contrasts can be examined as well (note: not all of the ensuing contrasts represent orthogonal comparisons, but they do address interesting theoretical questions). For example, if we wanted to test the effects of the order of symbolic and conceptual knowledge in relation to PS performance, we will compare PS outcomes for conditions 4-6 vs. 7-9. There may be a slight advantage for the latter group of conditions if early CK acquisition is, indeed, facilitative. To see if it is possible to teach PS before SK, we will test conditions 7-9 vs. 10-12. We suspect that conditions 7-9 will show better performance than 10-12. To examine the CK impact on PS would involve testing the conditions 1-3 vs. 4-6. We posit better PS performance as a function of having received previous CK instruction. PS should also show benefits from being preceded by CK and SK (regardless of order; 4-9 vs. 1-3 & 10-12).

Order × Aptitude Interaction. If we do not find main effects dues to order, we may find an aptitude-treatment interaction. For instance, individuals with higher inductive-reasoning skills (perhaps in conjunction with working memory capacity) may profit more from having CK presented *prior* to SK, while those with lower inductive skills (or even higher associative-learning skills) may optimally benefit by the reverse pedagogical approach (i.e., teaching SK first, followed by CK). Cognitive abilities will be assessed prior to the learning portion of the experiment.

Analogies Effect on CK Acquisition. Explicitly instructing conceptual knowledge via analogies may serve three purposes: (a) enhance the memorability of CK elements (over time, a *retention* issue), (b) facilitate the acquisition of related PS elements (a *learning time* issue), and (c) help learners solve related PS elements (in real-time, a *transfer* success issue). Learning outcome and efficiency measures are hypothesized to be positively affected by this approach. That is, a person's performance should improve. Moreover, once the *AHA!* experience occurs via the analog, then the learning rate should also be accelerated. To empirically test these issues, we will examine (across various periods of time--immediately and following a lag) the effects of different treatment

conditions on learning outcome and efficiency for each outcome type (SK, PS, CK) including various aptitude measures in the equation.

Analogies vs. No Analogies. Conditions 1, 4, 7, 10 (NA conditions) will be contrasted with 2, 5, 8, 11 (UA) and 3, 6, 9, 12 (MA) to examine any main effect due to analogies. Collapsing the two analogy conditions (UA and MA) and contrasting no analogies vs. any analogy will involve the following contrast: Conditions 1, 4, 7, 10 vs. 2, 3, 5, 6, 8, 9, 11, 12. We will also examine learning gains to see if the impact is equal across all outcome types. While we hypothesize that PS should see greater gains (particularly for low-ability Ss), there should also be a significant boost for the CK elements themselves.

Unified vs. Multiple Analogies Effect. A straightforward contrast between the UA and MA conditions (2, 5, 8, 11 vs. 3, 6, 9, 12) will demonstrate any main effects attributable to using single versus multiple analogies to teach conceptual knowledge. We expect that the multiple analogies will result in superior CK outcomes (and possibly PS outcomes as well) given they have been tailored to suit each concept. Alternatively, if the value of a unified framework makes the distinction among mean, median, and mode sharper, then learners receiving the unified analogy might be more likely to remember the procedural differences in calculating those values as well as the conditions under which each facet is most appropriate. That is, a unified analogy may encourage learners to compare and contrast the three measures of central tendency. Under this scenario then, we would expect superior performance on transfer problems that, for instance, ask the learner to pick and calculate which facet best allows them to answer a question about central tendency for a particular set of data. Learners with tailored analogies might be less able to make this sort of conceptual distinction since the separate analogies might not have encouraged learners to compare and contrast the facets of central tendency.

<u>Discussion</u>

Analogies appear to help the acquisition and retention of conceptual knowledge. A salient question then concerns the components and attributes of an effective, explanatory analogy. The

work of D. Gentner (e.g., Gentner, Rattermann, & Forbus, 1993) suggests that overlap in terms of relations among attributes and relations among relations might provide a useful metric in determining the "soundness" of an analogy (such as the soundness of the analogy between a seesaw and the mean).

In addition to the components of effective analogies, certain processes aid transfer. Efforts to make the mapping (and the functionality of the mapping) between base and target more explicit can help the learner to use the analogy more effectively. For instance, Catrambone and Holyoak (1989) found that instructions that led learners to focus on relevant mappings between a base and target during instruction also led them to spontaneously apply the analogy to the target, even after a delay between reading the base and solving the target. In general (across five studies), they found that transfer is enhanced when: (a) multiple examples are used rather than just a single one, (b) explicit, directive comparison instructions are given versus implicit (or no) directive instructions, and (c) more problem-solving experience is provided in contrast to passive observation.

<u>Future Work</u>

One concern in the present study is whether the seesaw analogy is the appropriate "unified" analogy to use. That is, perhaps the folding number line analogy might be more effective as an overall framework and thus, the results might be different if this analogy is used. Future studies can test this notion by systematically varying which analogy is used as the unified one and which ones are used as the tailored ones for each facet of a concept.

The results from the present study could, as a by-product, provide some information about whether learning and carrying out a procedural skill can also help one learn more conceptual information about an area. The argument is that as you become better at the skill, your attention is freed up and perhaps you can notice more about the task or domain than you would otherwise. This "noticing," besides providing information in and of itself, could support additional inferencing that you might not have the luxury of carrying out otherwise if you were struggling to do the procedure (as you were early in the procedure learning phase). While the present study

could support the idea that PS acquisition can make it easier to acquire CK, a more thorough test of this idea would a between-subjects design in which learners receive different degrees of practice on a procedural skill and then are tested for conceptual knowledge acquisition.

References

Anderson, J. R. (1993). Rules of the mind. Hillsdale, NJ: Erlbaum.

Catrambone, R., & Holyoak, K. J. (1989). Overcoming contextual limitations on problem-solving transfer. Journal of Experimental Psychology: Learning, Memory and Cognition, 15, 1147-1156.

Chi, M.T.H. (1993). Barriers to conceptual change in learning science concepts: A theoretical conjecture. Proceedings of the 15th Annual Conference of the Cognitive Science Society (pp. 312-317). Hillsdale, NJ: Erlbaum.

Gentner, D., & Gentner, D. (1983). Flowing waters and teeming crowds: Mental models of electricity. In D. Gentner & A.L. Stevens (Eds.), Mental models (pp. 99-129). Hillsdale, NJ: Erlbaum.

Gentner, D., Rattermann, M. J., & Forbus, K. D. (1993). The roles of similarity in transfer: Separating retrievability from inferential soundness. Cognitive Psychology, 25, 524-575.

Kieras, D., & Bovair, S. (1984). The role of a mental Model in learning to operate a device. Cognitive Science, 8, 255-273.

Shute, V.J., Gawlick-Grendell, L.A., Young, R.K., & Burnham, C.A. (in press). An experiential system for learning probability: Stat Lady description and evaluation. Instructional Science.

Shute, V. J. & Glaser, R. (1991). An intelligent tutoring system for exploring principles of economics. In R. E. Snow & D. Wiley (Eds.), Improving Inquiry in Social Science: A

Volume in Honor of Lee J. Cronbach (pp. 333-366). Hillsdale, NJ: Lawrence Erlbaum Associates.

Shute, V. J. (in press). SMART: Student Modeling Approach for Responsive Tutoring. To appear in: User Modeling and User-Adapted Interaction (Special issue on student modeling).

Sweller, J., & Cooper, G.A. (1985). The use of worked examples as a substitute for problem solving in learning algebra. Cognition and Instruction, 2 (1), 59-89.

Endnote

[1]Each CK condition has an associated subscript: NA = No Analogy, UA = Unified Analogy, MA = Multiple Analogies.

REACTOR ANALYSIS FOR TREATMENT OF WASTE WATER CONTAINING AQUEOUS FILM-FORMING FOAM (AFFF) WITH FENTON'S REAGENT

D.-T. Chin
Professor
Department of Chemical Engineering

Clarkson University
Potsdam, New York 13699-5705

Final Report for:
Summer Faculty Research Program
Armstrong Laboratory

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, Washington DC

August, 1995

REACTOR ANALYSIS FOR TREATMENT OF WASTE WATER CONTAINING AQUEOUS FILM-FORMING FOAM (AFFF) WITH FENTON'S REAGENT

D.-T. Chin
Professor
Department of Chemical Engineering
Clarkson University
Potsdam, New York 13699-5705

ABSTRACT

The U.S. Air Force uses Aqueous Film-Forming Form (AFFF) as a fire suppressant for aircraft storage and maintenance hangars and in fire training exercises. Following a fire-fighting event, the foam is washed-down with water. Release of AFFF to drainage systems has an adverse effect to environment and to municipal sewage treatment operation. AFFF may be decomposed by reacting with Fenton's reagent, which is an aqueous solution of H_2O_2 and a ferrous salt at pH 2 to 4.

This report describes a reactor analysis for treating AFFF-contaminated water with excess Fenton's reagent. Sixty thousand gallons (60,000 gal) of waste water containing 2% of AFFF is to be treated in 20 working days and the concentration of AFFF in effluent waste water is to be reduced to 0.02%. Mass and energy balance and heat transfer analysis are made to estimate reactor volume and cooling requirement over a range of an apparent reaction rate constant from 0.001 to 1.0 $M^{-1}min^{-1}$. Three types of reactors are considered: i) batch reactor; ii) continuous stirred tank reactor (CSTR); and iii) multiple tank cascade. The analysis leads to the following conclusions:

1. To carry out one batch operation per a 12-hour work day, a reaction rate constant of 0.01 $M^{-1}min^{-1}$ and a batch reactor of 6,400 gal in volume are needed for the treatment. The required heat transfer area of tank cooling coils and cooling water flow rate are: 155 ft^2 and 8.5 gal/min, respectively.
2. CSTR needs a large reactor volume because of low concentration of AFFF and H_2O_2 in the reactor. It is more economical to operate the CSTR at a large reaction rate constant of 0.1 $M^{-1}min^{-1}$. With this rate constant, the required reactor volume, cooling coil area and cooling water flow rate will be 5,500 gal, 35 ft^2 and 1.9 gal/min, respectively.
3. In multiple tank cascade, the total reactor volume is substantially reduced by connecting a number of small tank reactors in series. With a reaction rate constant of 0.01 $M^{-1}min^{-1}$, five tanks of 800 gal in volume are needed for continuous treatment of waste water. If the reaction rate constant is increased to 0.1 $M^{-1}min^{-1}$, only three tanks of 200 gal in volume are needed for the treatment. The required heat transfer area and cooling water flow rate in each tank are approximately 19 ft^2 and 1 gal/min, respectively.

REACTOR ANALYSIS FOR TREATMENT OF WASTE WATER CONTAINING AQUEOUS FILM FORMING FOAM (AFFF) WITH FENTON'S REAGENT

D.-T. Chin

INTRODUCTION

Aqueous Film-Forming Foam (AFFF) is a fire suppressant manufactured by the 3M and Ansol companies. The 3M FC-203CF LIGHT WATER brand AFFF consists of: 70% of water; 20% of butyl carbitol [or 2-(2-butoxyethoxy)-ethanol]; 5% of alkyl sulfate salts; 1 - 5% of amphoteric fluoroalkylamide derivative; 1% of triethanolamine; and 1% of perfluoroalkyl sulfonate salts (Material Safety Data Sheet). The U.S. Air Force uses a 3% AFFF aqueous solution in fire suppression systems for aircraft storage and maintenance hangars and in fire training exercises. Following a fire-fighting event, the foam is washed-down with water and discharged to a drainage system. AFFF contains hazardous chemicals, and release of AFFF-contaminated water has an adverse effect to environment and to the operation of municipal sewage and water treatment systems (Brittain 1995; EA Engineering 1994).

The work at the Air Force Armstrong Laboratory's Environics Directorate has demonstrated that AFFF may be destroyed by reacting with Fenton's reagent. Fenton's reagent is an aqueous solution of hydrogen peroxide and a ferrous salt, such as ferrous sulfate or ferrous chloride (Walling 1975). Under a slightly acidic condition of pH 2 to 4, this solution produces hydroxyl free radicals, OH• (Ingles 1972; Loizidou et al. 1993; Monhanty & Wei 1993; Walling 1975):

$$H_2O_2 + Fe^{2+} + H^+ \rightarrow OH\bullet + Fe^{3+} + H_2O \qquad (1)$$

When an organic compound, such as RH, is added to Fenton's reagent, it is oxidized by hydroxyl free radical and hydrogen peroxide via the following reactions:

$$RH + OH\bullet \rightarrow R\bullet + H_2O \qquad (2)$$

$$R\bullet + H_2O_2 + H^+ \rightarrow ROH + H_2O \qquad (3)$$

Reaction (2) and (3) are catalyzed by the presence of Fe^{2+} and Fe^{3+} in the solution. The product, ROH, may be further oxidized by Fenton's reagent. The rate constant for Reaction (2) between hydroxyl free radical and organic contaminants in drinking water is on the order of 10^9 to 10^{12} $M^{-1}min^{-1}$ (Haag and Yao 1992).

A large number of refractory (or non-biodegradable) organic compounds may be degraded by Fenton's reagent. These include: formaldehyde (Murphy et al. 1989); chloroform (Chen et al. 1990); carbon tetrachloride (Peyton et al. 1995); trichloroethylene (Carberry 1990); alcohols (Ingles 1972); citric

acid, triethanolamine (Pan et al. 1992); benzene, toluene, xylene and their derivatives (Lou & Lee 1995; Walling & Johnson 1975); phenol, cresol and other phenyl alcohols (Al-Hayek & Dore 1990; Castrantas & Gibilisco 1990; Land & Ellis 1982; Preis et al. 1994; Vella et al. 1992; Walling & Johnson 1975; Wang 1992); chlorinated phenols (Carberry 1990; Koyama et al. 1994; Lee & Carberry 1992; Potter & Roth 1993; Ruppert et al. 1994; Tyre et al. 1991; Watts et al. 1990 & 1993); nitrobenzene, nitrotoluene, nitrophenols (Lipczynska-Kochany 1991 & 1992; Mohanty & Wei 1993; Peyton et al. 1995); polychlorinated byphenyls (PCBs) (Koyama et al. 1994; Pignatello & Chapa 1994; Sedlack & Andren 1991 & 1994); dichlorophenoxyacetic acid (Sun & Pignatello 1993b & 1993c); and dichlorvos (Lu et al. 1994). The apparent first order reaction rate constants, k_A, evaluated from various literature sources for degradation of organic chemicals by Fenton's reagent are summarized in Table 1. The table also includes several values of an apparent first order rate constant, $k_{peroxide}$, for consumption of hydrogen peroxide. The value of k_A decreases with increasing molecular size of organic chemicals; it varies from 0.023 min^{-1} for formaldehyde to 0.0003 min^{-1} for hexadecane. Excess hydrogen peroxide is used in most studies. Use of ultra violet (UV) irradiation or ultra sonic wave improves the reaction rate by a factor of 2 to 4 as demonstrated by several studies in Table 1.

The Fenton process has been used to treat strong industrial waste waters (Chowdhury 1974; Fagen 1994; Tascioglu et al. 1993); solid particulates from a corn processing plant (Mukherjee & Levine 1992); textile waste water (Groff 1993); photographic processing waste solutions (Tsyzuki & Miyazaki 1994); and leachates from industrial landfills (Loizidou et al. 1993). Several workers reported the destruction of insecticides and herbicides, such as atrazine (Arnold et al. 1995), 2,4-D, 2,4,5-T and chlorophenoxyalkanoic acid (Sun & Pignatello 1993a; Pignatello & Sun 1993) in ground water by Fenton's reagent. Studies on removal of organic contaminants in soil by Fenton's reagent, including hexadecane, benzo[a]pyrene, pentachlorophenol, creosote, and explosives are described in several governmental reports (Watts & Stanton 1994; Watts et al. 1994; Pfeffer et al. 1985).

The purpose of this work is to perform a theoretical chemical reactor analysis for treating AFFF-contaminated waste water with Fenton's reagent. Sixty thousand gallons (60,000 gal) of waste water containing 2% of AFFF is to be treated in 20 working days. The requirement for reactor size, heat exchanger area and cooling water flow rate are evaluated for the process to be operated in batch and continuous mode.

ANALYSIS

Apparent Reaction Kinetics - To simplify the analysis, we shall use the following overall reaction to describe the decomposition of AFFF in water by Fenton's reagent:

$$AFFF + y\,H_2O_2 \rightarrow Products \tag{4}$$

where y is the apparent stoichiometric coefficient of H_2O_2. To account for the effect of H_2O_2 concentration on reaction rate, we assume that the reaction follows an apparent second order reaction of the type:

$$-r_{AFFF} = -\frac{d[AFFF]}{dt} = k\,[AFFF]\,[H_2O_2] \tag{5}$$

where r_{AFFF} is the reaction rate based on number of moles of AFFF reacted per unit time per unit liquid volume; t is the time; [AFFF] and $[H_2O_2]$ are the molal concentration of AFFF and H_2O_2, respectively; and k is an apparent second order reaction rate constant. Equation (5) leads to the following relationship between the concentrations of AFFF and H_2O_2 in the course of reaction:

$$\ln\frac{[H_2O_2]}{[AFFF]} = \ln\frac{[H_2O_2]_o}{[AFFF]_o} + ([H_2O_2]_o - y[AFFF]_o)\,k\,t \tag{6}$$

where $[AFFF]_o$ and $[H_2O_2]_o$ are the initial concentration of AFFF and H_2O_2. According to Equation (6), a plot of $\ln\{[H_2O_2]/[AFFF]\}$ versus t would give a straight line, and the value of apparent reaction rate constant, k, may be evaluated from the slope of the straight line if y is known. Figure 1 shows such a plot for decomposition of o-chlorophenol by Fenton's reagent. The data were taken from the work of Potter and Roth (1993), and a linear regressional fit of the data resulted in a k of 2.8 $M^{-1}min^{-1}$. Table 2 summarizes the values of k evaluated from the literature sources for some organic chemicals reacting with Fenton's reagent. The value of k ranges from 1.3 X 10^{-4} $M^{-1}min^{-1}$ for trifluralin to 5.4 $M^{-1}min^{-1}$ for 2,4-dinitrotoluene. The experimental conditions including initial concentration of organic species, $[A]_o$, initial molal ratio of H_2O_2 to the organic species, $[H_2O_2]_o/[A]_o$, initial molal ratio of H_2O_2 to iron salt, $[H_2O_2]_o/[Fe]_o$, pH and temperature, are given on the 2nd to 6th columns of the table. The apparent stoichiometric coefficient, y, of H_2O_2 as evaluated from available literature data is given on the 7th column of the table. It should be noted that y in the table is an average value over the experimental duration. The value of y is generally small at the beginning of an experiment, and increases with experimental time because of further reaction between the decomposition products and hydrogen peroxide.

Basis of Analysis - The present analysis is based on treating 60,000 gal of waste water containing 2% of AFFF with Fenton's reagent (EA Engineering 1994). The water is to be treated in 20 working days, and the concentration of AFFF in effluent waste water is to be reduced to 0.02%. Since butyl carbitol is the main organic ingredient in AFFF, we shall use it as the key component and designate its molal concentration as the concentration of AFFF in the analysis. Fifty percent (50%) hydrogen peroxide solution in the molal ratio of 70:1 to butyl carbitol will be charged to the reactor. The initial liquid mixture in the

reactor contains approximately 15 mM of butyl carbitol and 1050 mM of H_2O_2. Reaction (4) is exothermic and a cooling coil and cooling water will be needed to maintain the reactor temperature at a constant value. We assume that the initial temperature of waste water fed to the reactor is 25°C and the temperature of liquid in the reactor is not to exceed 40°C (104°F). The inlet and outlet cooling water temperature are assumed to be 20°C (68°F) and 32°C (90°F), respectively. The heat of reaction for complete degradation of butyl carbitol to carbon dioxide and water as calculated from its heat of formation at 25°C (DIPPR 1995) is -7270 kJ/mol. An overall heat transfer coefficient of 0.284 kW/(m^2C) for a tank cooling coil (Perry et al. 1969) is used for computing the heat transfer area requirement. Table 3 summarizes the base quantities used in the analysis.

Type of Reactors - Three types of reactors are considered in this study; they are: i) batch reactor; ii) continuous stirred tank reactor (CSTR); and iii) multiple tank cascade.

The flow diagram of a batch reactor is shown in Figure 2. The waste water containing AFFF is pumped from a waste water holding pond to the batch reactor where it is mixed with hydrogen peroxide and ferrous salt. The pH of reaction mixture is adjusted to a desired value by adding sulfuric acid. We shall let the reaction proceed until 99% of AFFF in the waste water is destroyed by Fenton's reagent. Cooling water will be pumped through a cooling coil in the reactor to remove heat of reaction and to control the liquid temperature at 40°C. The batch reactor will be operated on a 12 hour-per-day basis. This is equivalent to treat an average of 250 gal/h of waste water. At the completion of reaction, the waste water is neutralized to pH 6 to 8 by adding NaOH to the reactor. We assume that there will be two hours of preparation time between two successive batch operations. The rate of filling and discharging a batch reactor with waste water is assumed to be 40 gal/min.

The flow diagram for the continuous treatment of AFFF-contaminated waste water with Fenton's reagent in a CSTR is shown in Figure 3. The CSTR is to be operated on a 24 hour-per-day basis, and waste water together with Fenton's reagent and sulfuric acid is continuously fed into the reactor at a rate of 125 gal/h. The liquid inside reactor is well mixed such that the treated waste water exiting from the reactor may be assumed to have the same composition as that inside the reactor. A separate neutralizing tank is needed for adjusting the pH of effluent waste water to a neutral value of 6 to 8 before being released into a waste stream. With the exception of waste water volumetric flow rate, the design basis for CSTR is the same as that of a batch reactor as shown in Table 3. The advantage of CSTR is continuous operation; it has the drawback of needing a large reactor volume due to low reaction rate at low concentration of AFFF and H_2O_2 in the reactor.

Figure 4 shows the continuous process with a multiple tank cascade. In this scheme, a number of tank reactors are connected in series. The waste water together with Fenton's reagent and sulfuric acid is

continuously fed to Tank 1 at a rate of 125 gal/h. The effluent of Tank 1 is fed to Tank 2 whose effluent is then fed to Tank 3, and so on. In multiple tank cascade, only the last tank has a low reaction rate governed by the final effluent concentration. All the other tanks have a higher AFFF concentration, and a higher reaction rate. Thus for a given overall conversion of AFFF, the total reactor volume will be expected to be less than that of a single CSTR.

In the present study, mass and energy balance and heat transfer analysis are performed for the above reactors to determine the requirement of reactor volume, heat transfer area and cooling water flow rate for treating AFFF-contaminated waste water with Fenton's reagent. Calculation is made for a range of k from 0.001 to 1.0 $M^{-1}min^{-1}$ and y from 2.5 to 20. A safety factor of 2 is used for estimating reactor size and heat transfer area of cooling coils. The details of analytical procedures are given in Levenspiel (1962), Chen (1983), and Perry et al. (1969).

RESULTS AND DISCUSSION

Batch Reactor - Using the second order reaction kinetics of Equation (4) and (5), the change in AFFF and H_2O_2 concentration in batch reactor may be calculated. Figure 5 and 6 are a plot of AFFF and H_2O_2 concentration versus time, respectively, for k = 0.01 $M^{-1}min^{-1}$ and y = 2.5 to 20. The curves were calculated using the initial concentration of AFFF and hydrogen peroxide described in Table 3. The concentration of AFFF decreases logarithmically with the time; the rate of AFFF degradation increases with decreasing value of y. On the other hand, the concentration of hydrogen peroxide decreases at a faster rate with increasing value of y.

Figure 5 permits one to determine the reaction time required for the AFFF concentration to decrease to the desired level of $[AFFF]/[AFFF]_o = 0.01$. For a value of k of 0.01 $M^{-1}min^{-1}$, the reaction time varies from 7.5 h at y = 2.5 to 9.5 h at y = 20. Figure 7 shows a plot of reaction time versus reaction rate consntant, k, at y = 2.5 and 20 over a range of k from 0.001 to 1.0 $M^{-1}min^{-1}$. The reaction time decreases logarithmically with increasing value of k. The required batch reactor volume, V, may be estimated from the reaction time, t_{rxn}, time of preparation between two successive batch operations, t_{prep}, average volumetric treatment rate of waste water, Q_{waste}, and volumetric rate of charge/discharge of waste water to the reactor, $Q_{c/d}$, by the following equation:

$$V = \frac{S_f Q_{waste} (t_{rxn} + t_{prep})}{1 - (Q_{waste} / Q_{c/d})} \tag{7}$$

where S_f is a safety factor for design calculation. Figure 8 shows the predicted reactor volume versus reaction rate constant for two values of y of 2.5 and 20. At k = 0.01 $M^{-1}min^{-1}$ and y = 20, the required

batch reactor volume is 6,400 gal.

In batch reactor, the temperature of an exothermic reaction mixture would rise continually if heat is not removed from the reactor. This temperature rise may be estimated from energy balance. Figure 9 shows the adiabatic temperature rise versus time for k = 0.01 $M^{-1}min^{-1}$ and y = 2.5 and 20. The maximum temperature rise in the present reaction mixture is 26°C, and the effect of y on the rate of temperature rise is small. Most of the temperature rise occurs in the initial stage of time corresponding to 90% of conversion of AFFF in the reactor. This time may be evaluated from Figure 5 by letting $[AFFF]/[AFFF]_o$ = 0.1. During this time, cooling will be required to control the reactant temperature at a constant value. The rate of heat removal from the reactor within this time period can be evaluated from energy balance, and the heat transfer area of cooling coil and volumetric flow rate of cooling water may be calculated from heat transfer analysis. Figure 10 shows the required heat transfer area and volumetric cooling water rate as a function of reaction rate constant for y = 20. The effect of y on the size of heat exchanger and cooling water requirement is negligible. At a k of 0.01 $M^{-1}min^{-1}$, the required heat transfer area and cooling water flow rate are: 155 ft^2, and 8.5 gal/min, respectively.

Continuous Stirred Tank Reactor (CSTR) - The volume of CSTR required for reducing AFFF concentration from 2% to 0.02% (i.e., $[AFFF]/[AFFF]_o$ = 0.01) is given in Figure 11 as a function of k for y = 2.5 and 20. With a high molal ratio of H_2O_2 to AFFF in the feed (α = 70), the effect of y on reactor size is small. For k = 0.01 $M^{-1}min^{-1}$ and y = 20, the required reactor volume to treat 125 gal/h of waste water is 55,000 gal. This large reactor volume is primarily due to operating the reactor at a low concentration of AFFF. The volume of CSTR may be reduced if one can increase the reaction rate constant by adjusting the load level of ferrous catalyst in reaction mixture. At a k of 0.1 $M^{-1}min^{-1}$, the reactor volume is reduced to 5,500 gal as shown in Figure 11.

The rate of heat removal for the degradation of AFFF with Fenton's reagent may be obtained by the energy balance of CSTR. Since CSTR is operated at low concentration and low reaction rate, its cooling requirement is smaller than that of a batch reactor. The required heat transfer area and cooling water flow rate for CSTR are: 35 ft^2 and 1.9 gal/min, respectively; the two quantities are independent of CSTR volume and reaction rate constant.

Multiple Tank Cascade - The drawback of a large reactor volume in CSTR may be overcome by using a number of smaller tank reactors that are connected in series as shown in Figure 4. Figure 12 shows the calculated AFFF effluent concentration in the multiple tank cascade operation as a function of tank volume for k = 0.01 $M^{-1}min^{-1}$ and y = 20. It is seen that with a tank volume of 800 gal, five tanks are needed to reduce the AFFF concentration from 15 mM to 0.015 mM with Fenton's reagent. An energy balance of the 5-tank cascade reveals that maximum cooling requirement occurs at Tank 2 where a cooling coil of

19 ft^2 in heat transfer area and a cooling water flow rate of 1 gal/min are needed to maintain the liquid temperature in the tank at 40°C. If one can adjust the concentration of ferrous catalyst in reaction mixture such that the value of k is increased to 0.1 $M^{-1}min^{-1}$, Figure 13 shows that only three tanks in series, each with a volume of 200 gal, are needed to reduce the AFFF concentration from 15 mM to 0.15 mM. The maximum cooling requirement in the 3-tank cascade occurs in Tank 1 where the required heat exchanger area and cooling water flow rate are: 18 ft^2 and 1 gal/min, respectively.

CONCLUSIONS

An analysis has been made to estimate reactor volume and cooling requirements for treating 60,000 gal of AFFF-contaminated waste water with Fenton's reagent in 20 working days. The concentration of AFFF in the waste water is to be reduced from 2% to 0.02%. Three types of reactors are considered: i) batch reactor; ii) continuous stirred tank reactor (CSTR); and iii) multiple tank cascade. Calculation is made for a range of an apparent second order reaction rate constant of 0.001 to 1.0 $M^{-1}min^{-1}$. The analysis leads to the following conclusions:

1. To carry out one batch operation per a 12-hour work day, a reaction rate constant on the order of 0.01 $M^{-1}min^{-1}$ and a batch reactor of 6,400 gal in volume will be needed for the treatment. The required heat transfer area of tank cooling coils and cooling water flow rate are: 155 ft^2 and 8.5 gal/min, respectively.
2. CSTR needs a large reactor volume because of low concentration of AFFF and H_2O_2 in the reactor. It is more economical to operate the CSTR with a large reaction rate constant on the order of 0.1 $M^{-1}min^{-1}$. With this reaction rate constant, the required reactor volume, cooling coil area and cooling water flow rate are 5,500 gal, 35 ft^2 and 1.9 gal/min, respectively.
3. In multiple tank cascade, the total reactor volume is substantially reduced by connecting a number of small tank reactors in series. With a reaction rate constant of 0.01 $M^{-1}min^{-1}$, five tanks of 800 gal in volume are needed for continuous treatment of the waste water with Fenton's reagent. If the reaction rate constant is increased to 0.1 $M^{-1}min^{-1}$, only three tanks of 200 gal in volume are needed for the treatment. The required heat transfer area and cooling water flow rate in each tank are approximately 19 ft^2 and 1 gal/min, respectively.

REFERENCE

Al-Hayek, N. and Dore, M., Water Res. 24, 973, 1990.

Brittain, J., Military Firefighter, Issue 25, 7, May 1995.

Carberry, J.R., Superfund '90: Proc. 11th National Conference, pp. 826 - 830, Hazardous Materials Control Research Institute, Silver Spring, MD, 1990.

Castrantas, H.M. and Gibilisco, R.D., ACS Symp. Ser. 422, pp. 77 - 99, American Chemical Society, Washington, DC, 1990.

Chen, J.R., Xu, X.-W., Lee, A.S. and Yen, T.F., Environ. Technol., 11, 829, 1990.

Chen, N.H., Process Reactor Design, Allyn & Bacon, Inc., Boston, 1983.

Chowdhury, A.K., Ph.D. Dissertation, University of Denver, Denver, Co, 1974.

DIPPR, American Institute of Chemical Engineers, New York 1995.

EA Engineering, Science and Technology, Sparks, MD, Draft of Report 11206.99 to the U.S. Air Force Armstrong Laboratory and Andrews Air Base, October 1994.

Fagan, M.R., Environ. Protection., p. 45, September 1994.

Groff, K.A., Water Environ. Res., 65, 421, 1993.

Haag, W.R. and Yao, C.C.D., Environ. Sci. Technol., 26, 1005, 1992.

Ingles, D.L., Aus. J. Chem., 25, 87, 1972.

Koyama, O., Kamagata, Y. and Nakamura, K., Water Res., 28, 895, 1994.

Land J. and Ellis, J., Water Res., 16, 1139, 1982.

Lee, S.H., and Carberry, J.B., Water Enviorn. Res., 64, 682, 1992.

Levenspiel, O., Chemical Reaction Engineering, John Wiley & Sons, Inc., New York, 1962.

Lipczynska-Kochany, E., Environ. Technol., 12, 87, 1991.

Lipczynska-Kochany, E., Water Poll. Res. Canada, 27, 97, 1992.

Loizidou, M., Papadopoulos, A. and Kapetanios, E.G., J. Environ. Sci. Health, A28, 385, 1993.

Lou, JC. and Lee, S.S., Hazardous Waste & Hazardous Materials, 12, 185, 1995.

Lu, M.-C., Roam, G.-W., Chen, J.-N. and Huang, C.-P., Water Sci. Tech., 30, 29, 1994.

Material Safety Data Sheet, p.554, 3M Center, St. Paul, MN.

Mohanty, N.R. and Wei, I.W., Hazardous Waste & Hazardous Materials, 10, 171, 1993.

Mukherjee, S.R. and Levine, A.D., Water Sci. Tech., 26, 2289, 1992.

Murphy, A.P., Boegll, W.J., Price, M.K., Moody, C.D., Environ. Sci. Technol., 23, 166, 1989.

Pan, B.Y.K., Yen, T.F., Chen, J.R., U.S. Patent 5,139,679, 1992.

Perry, J.H., Chilton, C.H. and Kirkpatrick, S.D. (ed.), Chemical Engineers' Handbook, 4th ed., pp. 11-21 to 23, McGraw Hill Book Co., New York, 1969.

Peyton, G.R., Bell, O.J., Girin, E. and Lefaivre, M.H., Environ. Sci. Technol., 29, 1710, 1995.

Pfeffer, J.L., Kubarewicz, J.W., Congedo, R.P. and Shea, T.G., Final Report of Contract No. DAAK11-81-C-0089 to U.S. Army Toxic and Hazardous Materials Agency, Engineering Science, Fairfax, VA, 1985.

Pignatello, J.J. and Chapa, G., Environ. Toxicology & Chem., 13, 423, 1994.

Pignatello, J.J. and Sun, Y, ACS Symp. Series 518, pp. 77 - 84, American Chemical Society, Washington DC, 1993d.

Potter, F.J. and Roth J.A., Hazardous Waste & Hazardous Materials, 10, 151, 1993.

Preis, S., Kamenev, S. and Kallas, J. Environ. Technol., 15, 135, 1994.

Ruppert, G., Bauer, R. and Heisler, G., Chemosphere, 28, 1447, 1994.

Sedlak, D.L. and Andren, A.W., Environ. Sci. Technol., 25, 1419, 1991.

Sedlak, D.L. and Andren, A.W., Water Res., 28, 1207, 1994.

Sun, Y. and Pignatello, J.J., J. Agri. Food Chem., 41, 308, 1993a

Sun, Y. and Pignatello, J.J., J. Agri. Food Chem., 41, 1139, 1993b.

Sun, Y. and Pignatello, J.J., Environ. Sci. Technol., 27, 304, 1993c.

Tascioglu, S. Aydin, A. and Kok, E., Fresenius Environ. Bull., 2, 491, 1993.

Tsuzuki, H. and Miyazaki, H., Nihon Shashin Gakkaishi, 57, 240, 1994.

Tyre, B.W., Watts, R.J. and Miller G.C., J. Environ. Qual, 20, 832, 1991.

Vella, P.A., Munder, J., Patel, B. and Veronda, B., 47th Purdue Industrial Waste Conference Proceedings, pp. 73 - 82, Lewis Publishers, Inc., Chelsea, MI, 1992.

Walling, C., Accounts Chem. Res., 8, 125, 1975.

Walling, C. and Johnson, R.A., J. Am. Chem. Soc., 97, 363, 1975.

Wang, Y.T., Environ. Prog., 11, 210, 1992.

Wattts, R.J., Spencer, C.J. and Stanton, P.C., Report WA-RD 338.1, Wasghington State Department of Transportation, Olympia, WA, 1994.

Watts, R.J. and Stanton, P.C., Report WA-RD 337.1, Wasghington State Department of Transportation, Olympia, WA, 1994.

Watts, R.J., Udell, M.D., Rauch, P.A. and Leung, S.W., Hazardous Waste & Hazardous Materials, 7, 335, 1990.

Watts, R.J., Udell, M.D. and Monsen, R.M., Water Environ. Res., 65, 839, 1993.

Table 1: Apparent first order reaction rate constant for decomposition of organic contaminants in water or soil by Fenton's reagent.

Contaminant & Environment	Initial Contamin. Conc. (mM)	Initial H_2O_2 Conc. (mM)	Initial Conc. Fe-salt (mM)	k_A (min^{-1})	$k_{peroxide}$ (min^{-1})	Reference
formaldehyde, CH_2O, in water	16.7	100		0.027		Murphy et al. 1989
chloroform, $CHCl_3$, in water, pH 5.5, 55°C with ultrasound	0.67	33.5	0.36	0.0177		Chen et al. 1990
nitrobenzene, $C_6H_5NO_2$, in water, pH 3.3 to 6.9, room temperature	0.1	8	0.035	0.003 (0.0043 with UV)		Lipczynska-Kochany 1992
2-nitrophenol, 2-$C_6H_4OHNO_2$, in water, pH 3.3 to 4.5 room temperature	0.1	8	0.035	0.0072 (0.0124 with UV)		
4-nitrophenol, 4-$C_6H_4OHNO_2$, in water, pH 3.3 to 5.3, room temperature	0.1	8	0.035	0.014 (0.048 with UV)		
				0.013 (0.041 with UV)		Lipczynska-Kochany 1991
2,4-nitrophenol, 2,4-$C_6H_3OH(NO_2)_2$, in water pH 3.3 to 5, room temperature	0.1	8	0.035	0.02 (0.066 with UV)		Lipczynska-Kochany 1992
2,4-nitrotoluene, 2,4-$C_6H_3CH_3(NO_2)_2$, in water pH 3.5 to 5, 21°C	0.42	6.3	0.76	0.034		Mohanty & Wei 1993

Table 1: continued.

Contaminant & Environment	Initial Contamin. Conc. (mM)	Initial H_2O_2 Conc. (mM)	Initial Conc. Fe-salt (mM)	k_A (min^{-1})	$k_{peroxide}$ (min^{-1})	Reference
o-chlorophenol, o-C_6H_4OHCl, in water, pH 3.5, 21°C	4.5	63	0.09	0.15	0.013	Potter & Roth 1993
m-chlorophenol, m-C_6H_4OHCl, in water, pH 3.5, 21°C	5.9	65	0.09	0.11	0.018	
dichlorophenol, $C_6H_3OHCl_2$, in water, pH 3.5, 21°C	5.0	60	0.085	0.0041	0.00073	
pentachlorophenol, C_6OHCl_5, in soil, pH 2, room temperature	0.19	1940	8.8	0.04 to 0.015	0.0055 to 0.003	Watt et al. 1990
pentachlorophenol, C_6OHCl_5, in soil, pH 3, 20°C	3.8	3500	7.2	0.0042	0.00068	Tyre et al. 1991
trifluralin, $F_3C(NO_2)_2C_6H_2N(C_3H_7)_2$, in soil, pH 3, 20°C	3.0	3500	7.2	0.0014	0.00095	
hexadecane, $C_{16}H_{34}$, in soil, pH 3, 20°C	4.4	3500	7.2	0.0003	0.00065	
dieldrin, $C_{12}H_{10}OCl_6$, in soil, pH 3, 20°C	2.6	3500	7.2	0.0002	0.0011	
monochlorobiphenyl, $C_6H_5C_6H_4Cl$, in water, pH 3, 25°C	0.001	0.027	0.015	0.0098		Sedlak & Andren 1994
2,4-dichlorophenoxyacetic acid, $Cl_2C_6H_3OCH_2COOH$, with 0.1 M methanol in water, pH 2.8, 25°C	0.1	10	0.1	0.00039 (0.0024 with UV)		Sun & Pignatello 1993b
dichlorvos, $(CH_3O)_2P(O))CH{:}CCl_2$, in water with UV	0.23	1.1	0.1	0.029		Lu et al. 1994

Table 2: Second order reaction rate constants between some organic contaminants and Fenton's reagent.

Type contaminant & Environment	Initial Contamin. Conc. [A] (mM)	Initial $[H_2O_2]_o$ to $[A]_o$ Ratio	Initial $[H_2O_2]_o$ to $[Fe]_o$ Ratio	pH	Temp. (°C)	App. Stoich. Coeff. H_2O_2	k ($M^{-1}min^{-1}$)	Reference
o-chlorophenol, o-C_6H_4OHCl	4.5	14	690	3.5	21	3.7	2.8	Potter & Roth 1993
m-chlorophenol, m-C_6H_4OHCl	5.9	11	720	3.5	21	2.9	2.0	
p-chlorophenol, p-C_6H_4OHCl	5.0	13	700	3	21	2.5	0.96	
dichlorophenols, $C_6H_3OHCl_2$	5.0	12	710	3.5	21	2.5	0.071	
2,4-dinitrotoluene, $C_6H_3CH_3(NO_2)_2$	0.42	15	8.3	3.5 - 5	21		5.4	Mohanty & Wei 1993
pentachlorophenol, C_6Cl_5OH, in soil	0.19	10200	220	2	room		0.018 to 0.0063	Watt et al. 1990
pentachlorophenol, C_6Cl_5OH, in soil, pH 3, 20°C	3.8	930	490	3	20		0.001	Tyre et al. 1991
trifluralin $F_3C(NO_2)_2C_6H_2N(C_3H_7)_2$, in soil	3.0	1180	490	3	20		0.00013	
2,4-dichlorophenoxyacetic acid, with 0.1 M methanol in water	0.1	100	10	2.8	25		0.039	Sun & Pignatello 1993b
chloroform, $CHCl_3$	0.67	50	94	3.5	55		0.53	Chen et al. 1990

Table 3: Basis of reactor analysis for treating AFFF-contaminated waste water with Fenton's reagent.

Design Condition	Batch Reactor	CSTR & Multiple Tank Cascade
volumetric treatment rate of waste water, Q_{waste}	250 gal/h	125 gal/h
initial or feed concentration of AFFF (in term of butyl carbitol) in waste water, $[AFFF]_o$	15 mM	
final or effluent concentration of AFFF (in term of butyl carbitol) in waste water, [AFFF]	0.015 mM	
initial or feed concentration of H_2O_2, $[H_2O_2]_o$	1050 mM	
molal ratio of H_2O_2 to AFFF (or butyl carbitol) charged to reactor, α	70	
2nd order reaction rate constant, k	0.001 to 1.0 $M^{-1}min^{-1}$	
apparent stoichiometric coefficient of H_2O_2, y	2.5 to 20	
initial or feed temperature of waste water, T_o	25°C	
temperature of liquid in reactor, T	40°C	
inlet cooling water temperature, $T_{cw,in}$	20°C	
outlet cooling water temperature, $T_{cw,out}$	32°C	
heat of reaction per mole of AFFF (or butyl carbitol), ΔH	-7270 kJ/mol	
design heat transfer coefficient, U	0.284 kW/(m^2C)	
design safety factor for reactor volume & heat transfer area, S_f	2	
preparation time between 2 batch cycles, t_{prep}	2 h	
waste water charge & discharge pumping rate in batch operation, $Q_{c/d}$	40 gal/min	

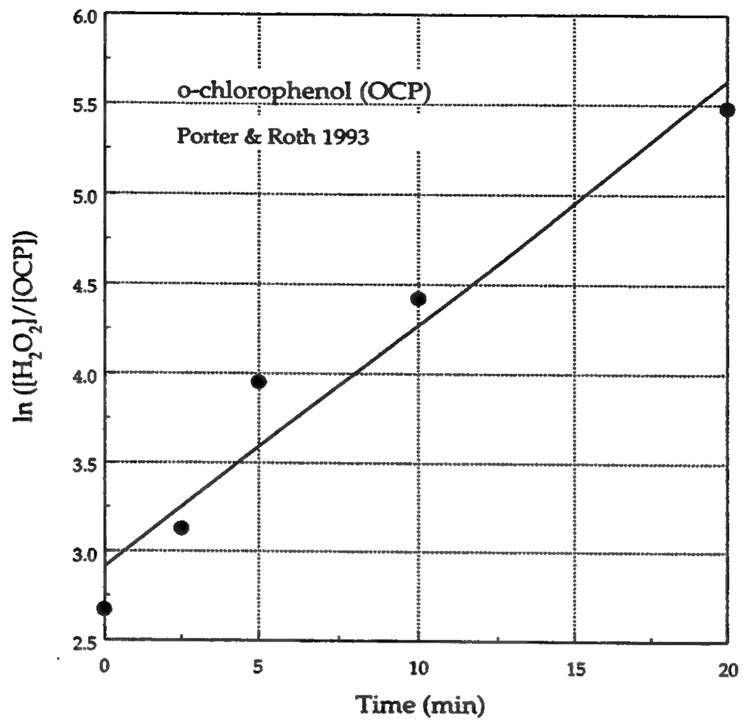


Figure 1: ln ($[H_2O_2]$/[OCP]) versus t for decomposition of o-chlorophenol (OCP) by Fenton's reagent. The data were taken from Potter and Roth (1993).

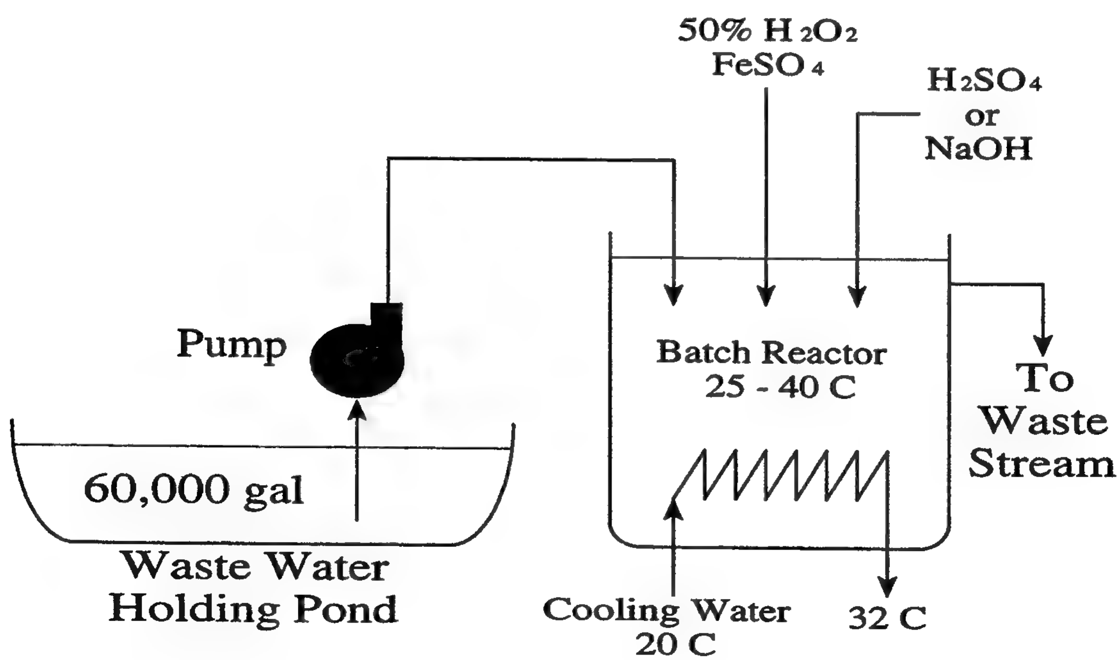


Figure 2: Flow diagram of a batch reactor for treating waste water containing AFFF.

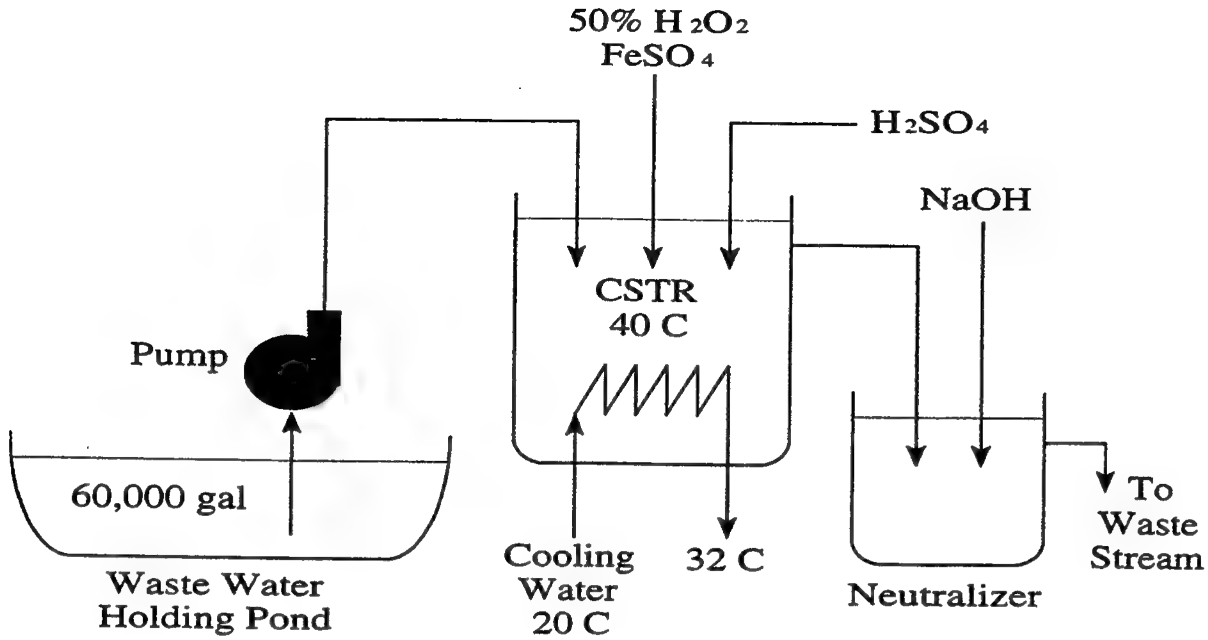


Figure 3: Flow diagram of a continuous stirred tank reactor for treating waste water containing AFFF.

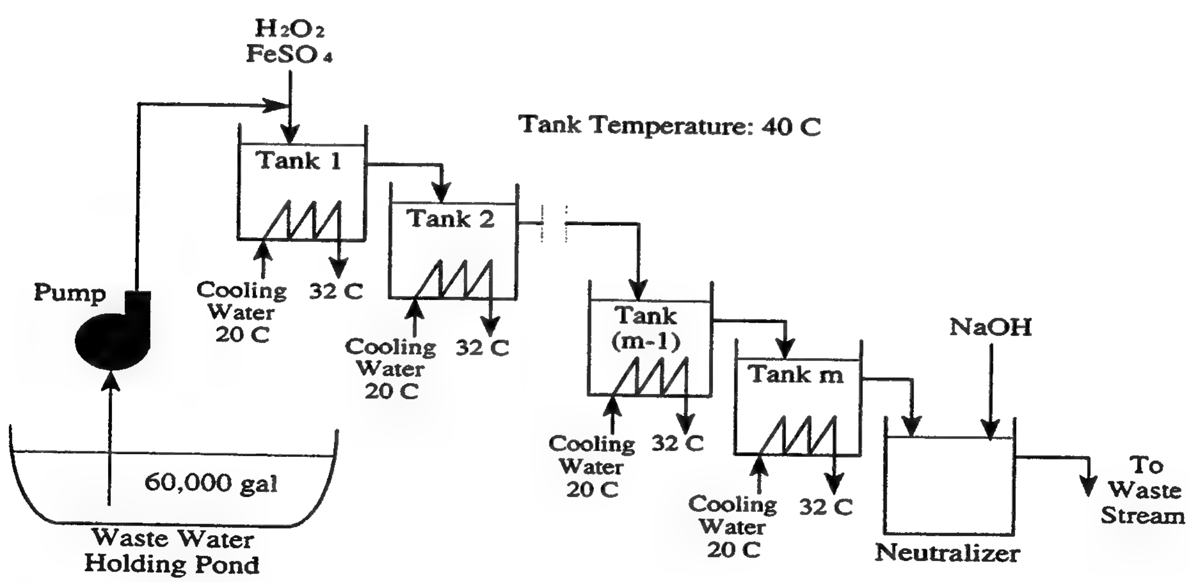


Figure 4: Flow diagram of multiple tank cascade for continuous treatment of AFFF-contaminated waste water with Fenton's reagent.

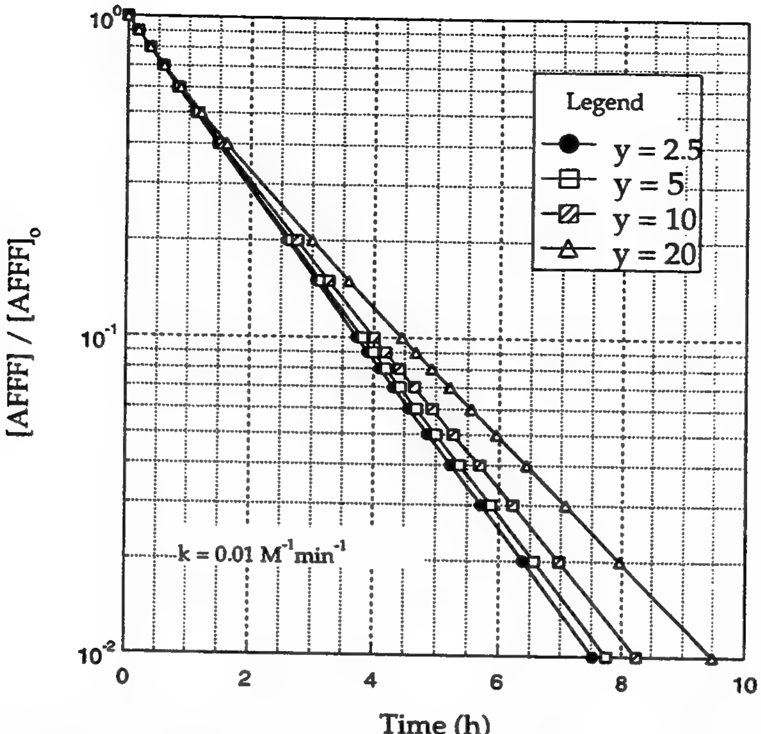


Figure 5: Concentration of AFFF versus time for a k of 0.01 $M^{-1}min^{-1}$. The initial concentration of AFFF and hydrogen peroxide are: 15 mM and 1050 mM, respectively.

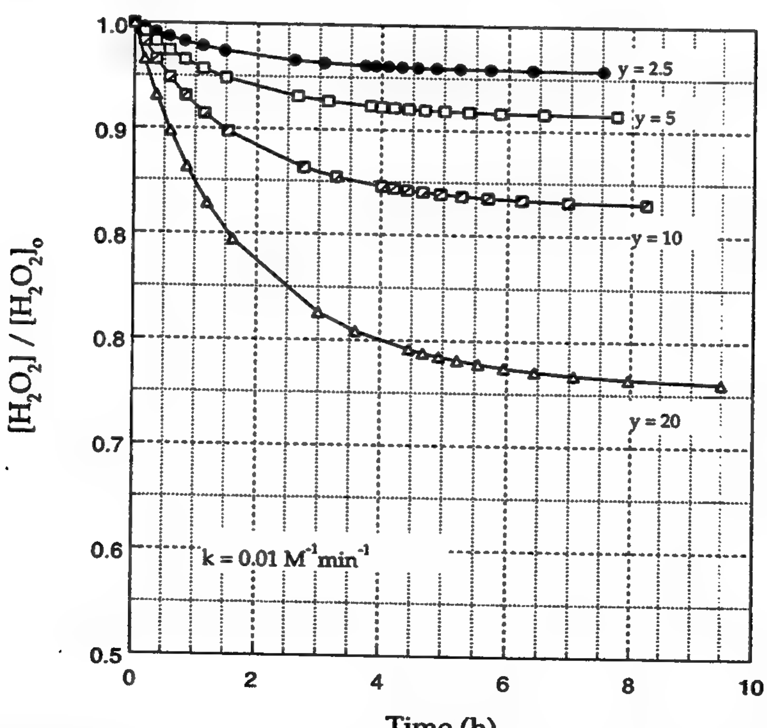


Figure 6: Concentration of hydrogen peroxide versus time for a k of 0.01 $M^{-1}min^{-1}$. The initial concentration of AFFF and H_2O_2 are: 15 mM and 1050 mM, respectively.

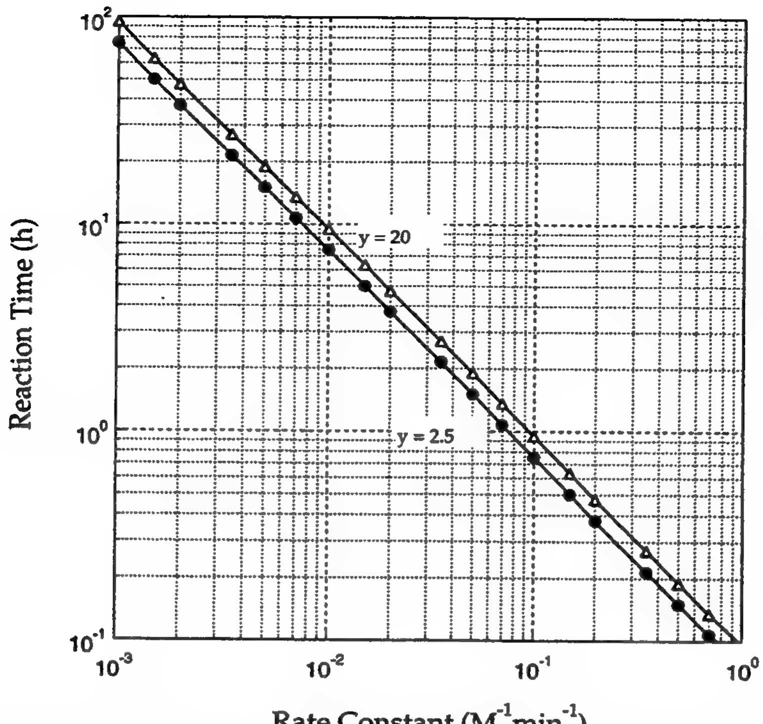


Figure 7: Reaction time versus k for reducing AFFF concentration from 15 mM to 0.15 mM in a batch reactor. The initial concentration of H_2O_2 is 1050 mM.

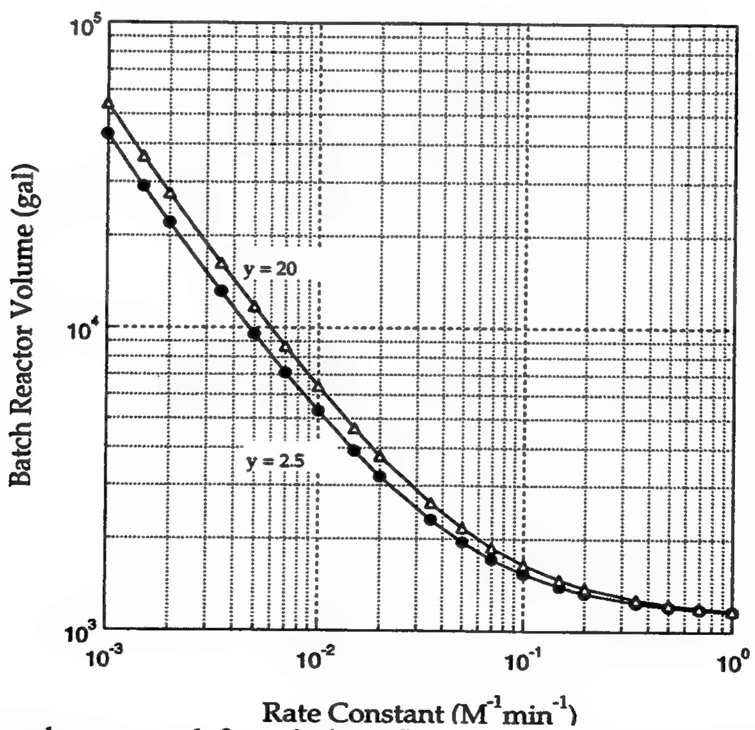


Figure 8: Reactor volume versus k for reducing AFFF concentration from 15 mM to 0.15 mM in a batch reactor. The initial H_2O_2 concentration is 1050 mM.

MULTI-OPERATOR PERFORMANCE AIDING

Gerald P. Chubb
Assistant Professor
Department of Aerospace Engineering, Applied Mechanics, and Aviation

The Ohio State University
164 W. 19th Avenue
Columbus, OH 43210-1100

Final Report for:
Summer Faculty Research Program
Armstrong Laboratory

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, DC

and

Armstrong Laboratory

September 1995

MULTI-OPERATOR PERFORMANCE AIDING

Gerald P. Chubb
Assistant Professor
Department of Aerospace Engineering, Applied Mechanics, and Aviation
The Ohio State University

Abstract

Performance aids can be viewed as any means of compensating for know human limitations in sensory, perceptual, cognitive, or motor performance capabilities. Such aids can assist or support the performance of problem solving, decision making, choice selection, or simple procedure execution. The aid can be built-into the designed system, supplied as a personal equipment item, provided as part of combat mission folder materials, or incorporated into published Technical Orders (e.g. checklists in the Flight Manual).

Multi-operator systems have a horizontal and/or vertical organizational structure. Systems with multiple crew members are a horizontally organized multi-operator system. Bombers, tankers, and transport aircraft well represent this class, but only some fighters have two crew members. Vertically organized multi-operator systems include interactions among cooperating, coordinating, or collaborative systems, such as between tanker and refueling aircraft, between flight lead and wingman, between bomber and command center, between AWACS and interceptor, etc.

Performance aids can assist operators in virtually any phase of the mission, from takeoff to landing. Examples for improving existing aids and proposed concepts for new aids are presented as possible candidates for further research. Emphasis is placed on adopting a three-tiered approach to test and evaluation of aid effectiveness: 1) performance predictions based upon theory and modeling, 2) empirical studies in laboratory and simulator exercises, and 3) cross-validation of models and products in well-instrumented and carefully designed flight tests.

MULTI-OPERATOR PERFORMANCE AIDING

Gerald P. Chubb

Introduction

Multi-operator systems are those in which more than one person is required in order to accomplish one or more assigned tasks efficiently or effectively. The operators may or may not be in close proximity. They may only be coupled by some form of communication, which allows them to interact in accomplishing their respective responsibilities. This interaction can be horizontal, as in the case where everyone is on some kind of local area network (LAN), which for this discussion could include an aircraft's intercommunication (intercom) channel. Alternately, it could be vertical, as in the case where the interaction occurs between the aircraft and some other system, each of which now become subsystems of a larger, more global system, interconnected in some form of wide area net (WAN), which in this case could include radio communication. This would include the aircraft's interactions with other aircraft, air traffic control facilities, or various command authorities (e.g the E-3A aircraft: Airborne Warning and Control System (AWACS) or the Army's equivalent: Joint Surveillance, Targeting, Attacking Radar (JSTARS)). Although relevant to overall system performance, our focus is not on the net or the communication system, but on the human operators and their performance.

There are cases where the activities an operator is expected to perform become sufficiently complicated, difficult, or lengthy that some form of performance aid is needed (or useful) to assist the operator overcome some natural human limitation, like short term memory or attention span, both of which are typically viewed as limited capacity human resources. Checklists are a common and typical example of such an aid. Procedural steps are identified to help eliminate errors of omission and commission in the execution of critical or lengthy task sequences. So-called decision support systems are another perhaps more exotic example of a performance aid.

Performance aids could be either built-into a system during its design, provided as a piece of personal equipment (much like helmets and parachutes), incorporated into Technical Orders (as checklists are in the Fight Manual), or provided as part of the mission materials (like maps are). With the advancements being made in expert systems technology, autonomous agents and other applications of artificial intelligence can also be viewed as advanced forms of performance aids. Because of the growing concern that "clumsy automation" can degrade rather than enhance performance (Wiener, 1988), it is important that aids be developed in such a fashion that their impacts on performance be predicted and validated prior to full-scale implementation. How that testing is accomplished can be as important as the development of the aiding concept itself.

With the increased difficulty and complexity of performing military missions, coupled with the increasing complexity of the technology being incorporated into the avionics of advanced weapons delivery systems, there appear to be many opportunities for aiding operators in performing their duties so system performance is not adversely affected but enhanced. Three problems have to be solved to achieve that objective: 1) identify areas where performance deficits exist which performance aids might alleviate, 2) synthesize one or more potentially effective aids, and 3) demonstrate that one or more of these solutions actually work. This summer's effort focused on trying to conceptualize the nature of this problem, possible approaches, and potential products which might be explored further.

Discussion of the Problem

While the global perspective is multi-operator performance aids, the local perspective is focused on performance aiding in manned strategic weapons delivery platforms: the B-1B and the B-2. It is therefore of some importance to identify the nature of the system-specific operating context and then to generalize that as much as possible to see where similar problems might occur in other operating contexts. Some problems will occur in both military and civilian operations, and the latter are certainly unclassified and amenable to study in academic research environments. Other problems only occur in the context of military operations, and some of those are quite sensitive, requiring access to classified information. Classified research studies are not typically suitable for masters theses or doctoral dissertation research. There was interest in identifying opportunities for creating products in each of these dual domains: military and civilian aviation. This precludes unnecessary duplication of effort, fosters cross validation of results (where possible and when needed), and clarifies what can or should be done by each of the participants in a collaborative research program.

One of the prominent unanswered questions is whether multi-operator performance aids in any way need to be designed differently than single-operator performance aids. Is there anything about the multi-operator interactions which requires an alternate design approach? That question implies looking carefully at those cases where performance aiding either occurs at the interface between crew members (cooperation, collaboration, and synchronization of activities) or it occurs in support of an individual preparing for such interchanges.

Methodology

The approach was largely conceptual and historical. Discussions with branch members and selected managers identified the need to look at what changes in system architecture and technological innovation suggest or motivate examination of performance aiding issues, identify potential products that

might address real or hypothesized performance deficiencies, and sketch out the steps needed to develop and test possible aiding concepts. The goal is to demonstrate the opportunities for a two-way technology transfer: 1) the incorporation of emerging human performance technologies into multi-operator performances aids, and 2) the export of models and methods derived from system-specific studies to other system research and development or operational testing efforts. Only limited bibliographic research was attempted.

Three other exercises were accomplished as part of this summer's effort: 1) review of four Mission Area Plans (MAPs) related to multi-operator systems, 2) field observation of a RED FLAG exercise at Nellis AFB, NV, and 3) review of selected unclassified B-2 documentation related to in-flight re-planning (IFR) performance aids. The last of these is relevant to studies being conducted by the Armstrong Laboratory. The particular RED FLAG exercise observed was multi-national. It included air drop as well as strategic and tactical bombing with fighter escort and defense suppression. Eight fighter pilots were interviewed, based upon a survey instrument prepared for an earlier study. Data reduction of the audio tapes has not been accomplished but is proposed. The review of the MAPs resulted in a separate document of some 51 pages which was delivered to Dr. June J. Skelly (AL/CFHI).

Conceptually, the search for performance aiding opportunities focused on elements of the mission and Oransanu's (1993) taxonomy of decision making: 1) simple choices -- (a) go/no-go and (b) condition-action; 2) resource allocation -- (a) objects or (b) time; and 3) problem solving -- (a) diagnostic or (b) creative. The mission was broken into typical phases: taxi and takeoff, climbout and cruise, aerial refueling, penetration, weapons delivery, and aircraft recovery. Search and rescue was added as a possibly relevant war-time support mission. Reconnaissance should be added as well, and the MAPs review (Chubb 1995a) certainly demonstrated a number of possible applications. In support of the Orasanu taxonomy, the production system paradigm was used, where situation awareness forms the front-end condition specification process and is comprised of two elements: 1) monitoring the current context, and 2) mental modeling to discern the future implications of present states. Actions are either pre-defined or some form of problem solving activity. Jensen, et al. (1994) and Chubb and Jensen (1995) provide recent reviews of literature related to aeronautical decision making. Also, Warner et al. (1995) provide an interesting examination of issues associated with predicting measures of system effectiveness from behavioral measures of performance in order to determine what factors have the largest impact on mission outcomes.

Much of the initial conceptualization was developed in a series of annotated briefing materials, using Microsoft POWERPOINT, which were delivered to Dr. June J. Skelly, AL/CFHI, Wright-Patterson AFB for her revision and use in preparing internal research program planning documents.

Results

This section summarizes some of the historical, RED FLAG exercise, technological, MAPs review, and methodological issues impacting multi-operator performance aid development and test. Recommendations are made for exploratory research and a three-tiered evaluation procedure is proposed: 1) constructive modeling and simulation, 2) laboratory and simulator studies, and 3) flight testing.

Historically, bombers were larger airframes, having larger crews. The B-52 had three, two-person teams: 1) pilot / co-pilot, 2) navigator / radar navigator, and 3) electronic warfare officer / gunner. The B-1 reduced this to a crew of four: 1) pilot and co-pilot, and 2) Offensive and Defensive Systems Operators (OSO / DSO). The B-2 has only two crewmembers: pilot and co-pilot. The mission is essentially the same; the speeds (therefore event pacing) have increased from about 360 knots (B-52) to about 540 knots (B-1B), a factor of 1.5 (50% faster); and targets and defensive threats have increased in number and complexity, as have the weapons used against them. All these factors can impact crew workload in one fashion or another. Automation has been a mixed blessing, often bringing new problems of its own, for example mode confusion (Sarter and Woods, 1992).

Air drop is similar to gravity weapons delivery, when performed at moderate altitudes. Some forms of air-drop require different tactics that require precision approaches, similar to touch-and-go landings, but without contacting the ground / runway. Since cargo aircraft are not typically equipped with self-protective systems, they are vulnerable to attack unless escorted by appropriately equipped fighters (or modernized to include self-protection systems (Electronic Counter-Measures (ECM), chaff, and flares).

Finally, while examples of multi-person performance aids for crew members is certainly of prime interest, multi-person aids can also be proposed for supportive operations, such as decision support systems (aids for command and control) and diagnostic aids for troubleshooting and repair of equipment, both on- and off-aircraft. Many examples of maintenance performance aids exist, some of which are now quite dated (e.g. Wilmot, et al., 1969 and Grieme, et al., 1969), but still not often applied.

The RED FLAG exercise is designed to provide practice for multi-element air combat operations. Observing the multi-national version of this exercise at Nellis Air Force Base this summer, it was apparent that any tactical situation display would need to classify objects in several different categories within friend-foe-neutral: 1) those ships in one's own formation (e.g. other fighters / bombers), 2) those ships working with us (escorts, SEAD, jammers, tankers, etc.), and 3) other own-forces ships. For the aggressor side, one would also like to separate bombers from fighter escorts and other types of support

aircraft, so target priorities could be clearly established. Both shape and color coding appear to be particularly well-suited to providing appropriate performance aids in this context (aircraft groupings with identification of aircraft type, possibly including expected weapons load/mix, at least for fighters/bombers).

Technologically, stealth and increased maneuverability have led to an airframe which is more difficult to fly and more demanding. To stay "hidden" the B-2 pilots must consider the radar reflectivity which comes with changes in aspect angle, and the time when this becomes most critical is exactly the time when they are most preoccupied by other concerns: weapons delivery. So technological advances in one arena often imply a need for multi-operator performance aids in an area where none were previously required. Electronic warfare continually evolves as countermeasures are developed and counter-counter measures are then explored to make designs more robust to countermeasures.

Methodologically, performance aids are a means to aid the operator in the actual execution of an activity. Conceptually, there is no reason to limit performance aids to on-the-job activity support. Methods could also be developed to provide performance aiding in at least two other contexts: planning and evaluation (pre- and post-flight respectively). These aids might be of a fundamentally different character. The first would provide opportunities for mission rehearsal using so-called "virtual reality" technology (Durlach and Mavor, 1995). The second would provide for mission reconstruction with the purpose of extracting "lessons learned" that can be fed back both to this team and to other teams preparing to execute similar missions sometime in the future. This is not a training aid, which would focus on skill development or proficiency maintenance, but truly a performance aid: how can / should operators adapt / adjust proficient skills to an evolving mission environment? As enemy tactics and strategies change, so must ours.

Twelve different areas were identified as candidates for developing multi-operator performance aids, some being improvements to existing aids, others being entirely new aids:

1. **Rejected takeoff (RTO) aids**: a graphic prediction of rotation and lift-off points as power is added and achieved, showing the effects (and trends in those effects) of engine spool-up and thrust development, a concept borrowed from AGATE (a NASA Langley Research Center program: Advanced General Aviation Transportation/Technology Experiments).

2. **Air Traffic Control (ATC) Coordination**: data link aids are proposed with Mode S transponders, but the display formatting and crew response modes need further research. The Close Air

Support (CAS) "9-line" information transfer, especially from the Forward Area Controller (FAC) side, needs to be examined and improved to reduce verbal communication overload and assure better accuracy of transmitting critical targeting data.

3. **Aerial Refueling Coordination**: could benefit from 4D navigation aids for rendezvous; and continuing attention should be given to reducing the effort required to attain and stay in contact-position behind and below the tanker throughout the refueling process.

4. **Message Authentication and Data Verification aids**: double-checking critical information and providing a graphic, interpretive portrayal of large sets of alpha-numeric information which need to be checked for accuracy. Some means is needed for making differences (between what is in computer memory versus what should be there) stand out.

5. **Terrain-Following / Avoidance aids**: low-level flight safeguards the aircraft from premature detection by enemy radar sites but it also risks controlled flight into the ground (CFIT); good displays are essential and should give pilots good perceptual set for what will next appear (weather, vertical obstructions (towers, ridges, buildings, etc.), enemy threats (missiles, anti-aircraft artillery, and possible small arms fire), etc.).

6. **Protective Systems aids**: monitoring, interpreting, and reacting correctly to both the electronic signal environment and visual sighting of Surface to Air Missile (SAM) launches is critical to survival and requires coordination among crew members. Doing the wrong thing (including the right thing at the wrong time) can be catastrophic.

7. **Navigation and Fix Taking aids**: could include new methods which include multi-point pattern matching versus the present single point crosshair placement methods. The human-system interface (HSI) techniques for this approach should be simple to implement in software, but some questions may need to be addressed in terms of Kalman filter updates in the avionics software.

8. **Target Recognition aids**: include compensation for aspect or look angle as well as for prior-strike damage effects which might have distorted the target image as well as its surrounds. Creating a "virtual reality" representation of the target in context would help, but locating friendlies is also important to preclude fratricide, especially in CAS missions.

9. **Weapons Release aids**: with strategic arms limitations, nuclear scenarios can benefit from warhead management by using damage assessment / strike tactics (actually, this is true for any weapon for which limited supplies are available); operators could use an aid for determining miss distance and damage levels of prior strikes.

10. **Diagnostic / Troubleshooting aids**: if an aircraft sustains battle damage, it will be important to isolate the nature of malfunctions and assess the capabilities and limitations of the aircraft and how that might affect post-strike recovery plans.

11. **Inflight Replanning and Recovery aids**: route re-planning in-flight could occur for many reasons (unexpected threat, time critical target assignment, or inflicted battle damage); another possible problem is trying to find a suitable post-strike recovery base if the pre-planned options are unachievable: the aircraft cannot fly that far or they have been destroyed by enemy attack(s).

12. **Search and Rescue aids**: are needed in those cases where the crew is forced to abandon their aircraft but have themselves survived. Many aids already exists, but with the Global Positioning Systems (GPS) new information is available, particularly if differential techniques are used to triangulate information from multiple receivers.

Mission Area Plan (MAP) Reviews (Chubb, 1995a) reinforced the need for performance aiding in three particular areas: 1) finding targets, 2) assessing damage, and 3) re-planning after unexpected events. However, several other opportunities also became apparent:

1. **Drug Trafficker Intercept**: F-16's are equipped with both an Advanced IFF (Identification Friend or Foe) system and an external ID (identification) spotlight for visual night operations. Directing a multi-ship operation in search of general aviation (GA) aircraft could benefit from having a search strategy and implementation aid that assures complete coverage of the sweep area. Once an intruder is found, tracking and hand-off to other aircraft may be required. By analogy, this bears some resemblance to theater missile defense (TMD).

2. **Decision Support Aids for Drug Trafficker Detection and Tracking**: The Regional and Sectional Operations Control Centers (ROCC/SOCC) have an Automated Air Movement Data System (AAMDS) to acquire Federal Aviation Administration (FAA) flight plan information. Detecting intruders means finding aircraft which have not filed a flight plan and separating normal Visual Flight Rule (VFR) traffic from possible smuggling or illegal entry operations.

Automated decision aids might be quite helpful here.

3. **Integration of Multi-source / Multi-spectral Data:** Surveillance and intelligence products arrive from many different sources. They have to be interpreted to arrive at a composite understanding of enemy order of battle and possible intent. What can "deceive" one kind of imagining sensor may not confuse others, so cross comparisons of imagery often reveal important details. Aids should facilitate the detection, identification, and classification of militarily significant objects: providing input for Air Tasking Order generation and redirection of strikes in progress (especially time critical targets TCT, like TMD).

4. **Unmanned Aerial Vehicles (UAV) Control Aids:** Previous Armstrong Laboratory studies (both empirical (e.g., Aume and Mills, 1981) man-in-the-loop (MIL) and Systems Analysis of Integrated Networks of Tasks (SAINT) constructive modeling and simulation (M&S) efforts (Duket, et al., 1976)) of Remotely Piloted Vehicle (RPV) Drone Control Facility (DCF) are increasingly relevant to future concepts of operation. Aids to facilitate multi-vehicle launch, monitoring (flight-following), employment, and recovery may be quite valuable.

5. **Flying Efficiency Aids:** Many aircraft are being retrofitted with more efficient engines, and the GPS integration affords precision navigation. With suitable weather information (upper atmosphere winds and air density), it is possible to optimize a flight paths for minimum fuel consumption. This is a performance aid that could pay for itself!

6. **Intra- and Inter-Flight Coordination of Information:** More efficient means of "passing" information to others is needed, both in the context of joint operations, and in the context of sequential operations. Lessons learned about a target area, unexpected threat systems, or destroyed threat systems can be very important to aircraft which will be entering an area at a later time. Aids for communicating that data must be accurate as well as quick. Iconic battle management displays appear attractive here.

7. **Interpretive ROE Aids:** In many conflicts, the rules of engagement (RoE) include restrictions on where one can fly and when one can shoot (air or ground targets). Since there are no "sign-posts" in the sky, performance aids which show where everyone is relative to artificial boundaries would clearly be helpful. However, even more important is knowing what one can shoot (safely) and when "approval" is first required. Again, some form of iconic may help sort out permissible from inadmissible potential targets. The problem is to keep the display simple and uncluttered while giving valid information.

8. **Ambiguity Reduction Aids:** Future conflicts anticipate encountering regional forces having a mix of weapons (some from the former Soviet Union, some from allies, and maybe even some of our own). This makes identification of friendly, neutral, and enemy forces all the more difficult. An aid for sorting out who's who is even harder when both sides have the same "electronic signature." One solution could be a joint use of individual soldier-GPS locator and some form of inexpensive telemetry / transmitter / transponder so the pilot can at least identify "own troops" unambiguously.

9. **Aids for Theater Missile Defense Commitments:** The intent to automate (with human override) the whole process of attacking SCUD-type weaponry, in part by using laser technology, presents formidable risks to aircrews in the area of the ABL platform. Also, TCTs present a challenge to all participants. Keeping out of each other's way may be as important as attacking quickly. Determining who has the authority to fire when multiple players are involved will take some kind of coordination / communication aiding scheme to assure everyone knows what's planned and their role.

10. **Virtual Target Area Pre-view:** In addition to having good intelligence inputs, it would also be nice to include consideration of weather effects: reduced ceilings or visibility and changing winds' effects on obscuring smoke. This is an ideal application for virtual reality and synthetic imagery, where pilots could "rehearse" the ingress briefly during the quieter portions of a mission, like while on Combat Air Patrol (CAP).

11. **Mission Monitoring Aids:** It is difficult to watch and pick out significant events when observing multiple participants in a highly dynamic environment: watching one precludes watching others. Aids to know who to watch at a particular time would be helpful: those deviating from plan might be one criterion. Those "under attack" would surely be another. Finally, anything about which action is impossible is irrelevant: focus on actions which can be aided. Such criteria "filter" what needs attention from what can be temporarily ignored.

12. **Target Base Management Aids:** While there are pre-defined target priorities, multi-target strike operations complicate flight management by creating uncertainty. Pilots can (and do) eliminate targets opportunistically (e.g. when an unexpected threat system "comes up" and is attacked by SEAD aircraft) and may fail to eliminate certain other assigned targets. Collecting that information and getting it back through intelligence channels to air tasking planners creates a massive amount of data. Correlation aids would help assure targets are not missed, prematurely dropped, or re-assigned when they do not need to be.

It is clear that any in-flight re-planning aid should consider on-going upgrades to the Air Force Mission Support System (AFMSS) and work directed toward strategic weapons delivery platforms (B-1 and B-2) should explore a similar effort being done for the F-15E Strike Eagle. Moreover, the RED FLAG exercises appear to well-represent the kind of environment within which this kind of aid would have to function if it became operational. In this context, a re-planner cannot restrict itself to a single aircraft: it must have data about the plans (or at least the intentions (Miller, Galanter, and Pribram, 1960): unfinished part of the plan) of other participants. The other two applications areas (finding targets and assessing damage) are inter-related, so an elaboration of possible aiding concepts is presented here for stimulating further discussion and research planning.

Pilots in a recent study (Chubb, 1995b) indicated finding the target was always a problem, whether in air-to-air combat or in air-to-ground weapons delivery: making the transition from inside to outside the cockpit, principally because "things look different." Real Time Intelligence in the Cockpit (RTIC) will not solve this problem unless the information is presented to the pilot in a meaningful fashion that facilitates the development of not only target cueing but an appropriate perceptual set (expectation). Since reconnaissance data are collected from the perspective of the collector (aircraft, satellite, or unmanned air vehicle), direct presentation will not "aid" the pilot. The target information needs to be transformed, perhaps even synthesized, to match the perspective of the particular pilot and the approach into the target area (heading, altitude, sensor depression angle, field of view, and display scale factor).

The appropriate perspective will depend upon the weapon, the tactics required to deliver that weapon, and the changing local conditions which may affect the run-in to the target. For example, if the defenses have changed position or been reinforced, then the approach into the target area may need to be changed to better assure survival of the delivery vehicle into the release point or to better set up the path of egress out of the target area. This could radically change the perspective view of the target from what was anticipated during mission briefing. Consequently, how the target information should be presented to the pilot may be time-sensitive: it may need to be changed to better equip the pilot to anticipate what will probably be observed in the target area.

This target intelligence information might be most appropriately presented tachistoscopicly in the imaging display. The principal problem in the case of air-to-ground weapon delivery is detecting, identifying, and recognizing the target, being sure it is the intended target and not an element of friendly forces (own or allies). What complicates the problem is weather: low (overcast) ceilings and restricted visibility reduce the amount of time available to look for, find, recognize, and attack the target. The operator needs to be cued where and when to look, but to be effective, the pilot must have an appropriate

expectation for what the target should look like, especially if affected by camouflage, concealment (e.g. partially hidden by foliage), or some other form of deception (e.g., decoys).

A second aspect of this problem is doing the bomb damage assessment (BDA) after a strike in order to determine the level of damage inflicted. The first problem is interpreting the collected imagery, comparing it with pre-strike imagery, which again, may have been obtained from an alternate perspective. The second problem is making the judgment itself: was the target sufficiently damaged to declare it as a catastrophic kill, or was it left sufficiently intact that another strike is warranted? This judgment may rest upon the comparison of pre- and post-strike imagery, which may be affected by time-of day (e.g. sun angle and shadow), seasonal, or other sensor effects (blooming, edge enhancements, absorption, scattering, cardinal point effects, glint, etc.). Adjustments to the imagery can modify one, the other, or both images in order to facilitate the comparison being made. While such comparisons can be viewed as static, the availability of multiple imagery (collected as a series of images over time) might be re-played in compressed time, real-time, or lapsed time and heighten certain differences / contrasts between images.

While image interpretation in a static planning environment is useful in supporting the generation of Air Tasking Orders (ATO), the combination of the interpretation and delivery decisions occurs in BDA/ Strike (S) sorties. These typically occur in situations where weapons or warhead management is of concern because of a limit in the number of weapons available of some particular type. Desiring to preserve scarce assets, a BDA/S crew must not only find and recognize the target, they must interpret, in real-time, the level of damage inflicted, and then decide whether to release or withhold the weapon for this particular target. This simply takes the intelligence problem and puts it in the cockpit along with the weapon delivery decision. It takes a complex and difficult task and makes it even more complicated and difficult. Clearly some form of performance aiding is probably needed.

One possible approach is to create, from real or synthesized imagery, a set of target images which represent different levels of inflicted damage, as comparative standards for making the damage assessment judgment. The old imagery of the target (before any strike or after the previous strike) could first be compared with a sequenced set of images (none to total destruction) to establish a baseline for the comparison. Second, the post-strike imagery could then be compared to the pre-strike imagery to determine whether a significant change had occurred. Finally, for significant changes, the question is whether the destruction was severe enough or not. A first comparison could match the target image with the standards for "best match" and then compare the target and standard to the criterion established for this target. Clearly, other comparison schemes are possible and worth exploring.

CAS is another area in which performance aids are needed to facilitate the transfer of the so-called "9-line" information from FAC to pilot. In a fluid combat environment, portions of this information are difficult to communicate unambiguously. There is a need for a graphic-oriented cartographic-based product that would allow the pilot to point to an object on the FAC's display and confirm they are both talking about the same target (or non-target). Such performance aids could reduce the risk of fratricide as well as increase sortie effectiveness. The other problem such aids could resolve is to give the pilot latitudes and longitudes for all points the FAC is identifying in Universal Transverse Mercator (UTM) units. This eliminates the need to even deal with UTMs in the cockpit environment.

Research, design, and development will be required in each case to take these concepts from preliminary ideas to proven applications. Preliminary research should focus on validating needs. That means determining the degree of system performance decrement or performance risk, and how performance aiding can be made most effective: what aspects of behavior need to be aided: 1) sensory / perceptual processes, 2) cognitive processing of perceived information (to determine the situation or operating condition), 3) decision making or problem solving activities (to formulate the appropriate course of action), or 4) action implementation and feedback (assuring the chosen goal is actually achieved, possibly detecting new problems).

System performance decrements are a quantifiable negative change in mission or combat effectiveness as a result of some inefficiency or error in crew behavior, e.g. increased weapons delivery miss distance resulting from increased reaction times and less accurate target recognition due to excessive crew fatigue. Since operators tend to compensate for their perceived degradation, system performance measures tend to be highly robust: not showing any decline even when the human capabilities are clearly degraded. People find ways to compensate in order to sustain system performance. However, there is still the risk that performance degradation will occur: it is less likely that the operator can absorb additional increases in load without serious consequences, because the margin left for further compensation has been reduced. Therefore, the concept of performance risk assessment potentially offers a more sensitive measure than performance decrement (Chubb, 1992).

Design of the aid itself depends upon isolating the behavioral deficiency, clarifying the nature of the problem, proposing innovative solutions, and then developing the specific implementation that operationalizes the performance aiding concept. In some cases, that may mean emulating a system design change, in hardware, software, or both. In other cases, it may mean developing something crew members use in conjunction with the system: 1) a piece of personal equipment (like calculators, slide rules, maps and protractors, etc.), 2) an addition to current materials (Technical Orders, combat mission materials,

checklists, etc.), or 3) an alternate procedure or operating technique, perhaps simple enough to learn without formal training (but not excluding some form of training).

Exploratory research efforts should focus on the development of theories and models which support: 1) isolating the performance deficiency to particular behavioral limitations, and 2) evaluation of aiding effectiveness. Constructive modeling and simulation of behavior in the context of crew duties should be pursued to estimate the possible system / mission payoffs of using performance aids. Such models can also be used for sensitivity testing to assist in the design of suitable scenarios for laboratory, simulator, and flight test studies. As discussed in Chubb and Hoyland (1989), prior studies (e.g., Bachert, et al., 1983) used this approach successfully, but it has not seen widespread use. Exploratory research should also address: 1) the mechanics of dynamic decision making (choices and selections in evolving environments), 2) attentional shifts and concentration in the presence of multiple distracters, and 3) the development and validation of good mental models of the situation. Many opportunities exist for developing problem solving aids, if not for real-time implementation, then at least in the pre- and post-flight contexts (planning and debriefing: feeding back and exploiting lessons learned from combat experience).

A three-tiered approach for evaluation of concepts (proposed alternative performance aid designs) is recommended: 1) analytic and predictive assessment of potential behavioral improvements (expected results), 2) laboratory and simulator studies of effectiveness (part and whole task validation of models and their predictions), and 3) flight testing (final product verification and validation of simulation studies). Since each of these becomes more expensive as well as more realistic, it is important to do them in sequence so defects are discovered and refinements made early, where re-testing is least expensive.

Analytic and predictive assessment using constructive modeling and simulation is contrasted with man-in-the-loop simulation, where a person interacts with some simulator or training device. Constructive models have often been either discrete event, network, or continuous system simulations. Relatively few combined simulation models have been developed, exploiting all three of these approaches. Although techniques are available for doing so (e.g. Pritsker, 1995), combined models are rare, because most engineers use one or the other approach, typically to the exclusion of the alternative approaches. Moreover, most of the military modeling has not used techniques like SLAM, but have developed one-of-a-kind models, most of which incorporate few if any considerations of human factors, with the exception of TAC BRAWLER (Bent, et al., 1991). Efforts to model systems, starting from the cognitive perspective of the operator, are only in the exploratory stage.

SOAR-based models (Laird, et al., 1986) tend to focus on the cognitive processing of internalized information, assuming the sensory / perceptual and motor aspects of performance are not a major concern. While that serves well in studies of cognitive psychology, it is not always adequate in the context of human engineering. As GOMS modeling demonstrated (Card, Moran, and Newell, 1983), micro-level descriptions tend to need empirical recalibration when applied to macro-level behavior predictions. Human-engineering models ought to place more emphasis on the environmental impacts and the crew-system interaction, issues which have typically been ignored in the academic models of cognitive behavior. Moreover, the impact of performance degradation factors is an issue, both in modeling and experimentation: how to estimate the impact of non-ideal operating environments -- first, on human behavior, and second, on system performance and mission success. Techniques like SAINT (Chubb, 1978) have proven successful in this context (Seifert and Chubb, 1973), if suitable data or estimates can be provided (which is admittedly difficult at best).

There has been a tendency for modelers to concentrate either on: 1) areas of technical competency, or 2) problems of current interest. As a result, most models are focused on issues which may not be of interest or relevance in others' studies. Models by behaviorists short-change systems and mission fidelity. System and mission modelers short-change behavioral impacts of human factors. Clearly, a more balanced approach might help each community better understand the relevance of the others' concerns and representational problems.

<u>Laboratory and simulator of effectiveness</u> can validate constructive modeling predictions but such studies require interpreting a sequential stream of behavioral data composed of one or more elements of: 1) audio-visual recordings, and 2) computer-measured data. The computer measured data are typically of two sorts: a) sampled time-trajectories of continuously changing variables (altitude, airspeed, etc.), and time-tagged discrete events (changes in switch settings, mode changes, etc.) Also, the empirical studies can be done with: 1) well-controlled laboratory tasks, 2) more realistic but artificially contrived "synthetic tasks" (Chubb, 1991), or 3) with some form of simulation: a) a ground-based simulator, or b) airborne flight-test. Laboratory tasks typically focus on a single person doing a single task. Synthetic tasks try to add realism by including some form of time-sharing among multiple concurrent tasks. However, top-down filtering of real-system details can lead to a synthetic task which no longer has the difficulty or complexity of the real task. Predictions of real system behavior from such studies then becomes difficult, and validation in simulations studies is surely recommended

Since system simulation studies often involve multiple operators each of whom may be performing multiple, concurrent tasks, logically separating an observed sequence of actual activities into its component chains of behavior is neither obvious or simple. SHAPA provides a tool for assisting this

process (Sanderson, 1994). It does not provide a complete theory for implementation. AKADAM and TASK (McNeese, et al., 1995) are directed toward understanding better how individuals choose to represent their mental models of their environment and task requirements (see also Zaff, et al., 1993), which in turn both enables and limits their awareness of the situation and their ability to formulate and execute appropriate courses of action, including coordination, collaboration and synchronization of activities. A good model of tasks is needed to interpret real-time, sequential, behavioral data steams.

Coordination is viewed as an informative means of cooperating when otherwise independent on-going activities could interact (affect one another). Collaboration is seen as an activity where no one person can do the task; multiple people must each do their part in order to achieve a common goal. Synchronization is seen as independent activity in which one or the other person's goal (possibly both) depend on another person either starting or completing a particular activity at a specific time. Synchronizing completions is inherently more difficult than initiation synchronization because of the need to predict an uncertain process: performance duration.

Since any interaction is inherently a communication process through some medium, the nature of human communication becomes a central issue. Efficient communication occurs against a cultural backdrop, where both speaker and listener are either a part of a common culture, or the communication is cross-cultural. Sathe (1985) offers a useful definition of culture: shared assumptions about beliefs and values. Beliefs are strongly associated with situation assessments, and values are strongly associated with attaching importance to alternative courses of action. Conflicts among multi-operator teams (about goals, plans, intentions, situation assessments, courses of action, expected outcomes, status of implementation, actual outcomes, etc.) will need to be resolved efficiently and require communication to do so. Certainly, some form of performance aid may be required or desirable in the communication domain, as well as in the sensory / perceptual, cognitive ,decision, and action domains.

Simulation studies serve to impose realistic levels of event pacing and crew workload in order to discern how well crews perform. Clearly, this can be done with and without the proposed performance aid, so a subject, at least in theory, can serve as their own control, using current practices as a baseline condition and the use of one or more performance aids as treatment condition(s). The scenario must itself be constructed in such a fashion that it imposes the problem in a natural (normal / realistic) fashion and evokes the behavior that needs to be aided so the adequacy of the response can be observed and measured.

Scenario construction is not a trivial task. Unobtrusive insertion of suitable "trigger" events and unobtrusive recording of crew behavior are not easily achieved without considerable planning. Artifacts are easily created unless anticipated and appropriately controlled. This is essential to assure minimal

confounding in the experimental results. Simulators are more complicated than laboratory tasks, and control is usually more difficult to maintain. What constitutes adequate (still incomplete) instrumentation of a simulator is also at issue and depends upon what measures are needed. Not all simulators or flight training devices (defined in Advisory Circular (AC) 120-45A: Airplane Flight Training Device Qualification, Federal Aviation Administration, Washington, DC) have been instrumented to collect performance data.

Flight testing is even more difficult since aircraft are not instrumented to collect crew behavior data and one has little or no control over in-flight contingencies: every flight is essentially unique in some sense, never to be truly replicated, only approximated. Getting an adequate means for observing and measuring what crews do and when they do it is important. Also, it is important to be able to compare the time of behavioral events with the time of system and mission events, so synchronization of data records is critical. TEST PAES (Smith, et al., 1994) provides a tool to help view flight test data, reduce it to meaningful measures, perform suitable analyses and comparisons, and interpret results. It also provides tools for managing a large flight test program.

Conclusion

Many opportunities exist for performance aiding in multi-operator systems, only some of which will enjoy dual application in civilian aviation. Target related performance aiding concepts clearly apply only to military aviation. However, in-flight re-planning potentially applies in civilian and general aviation as well. A three-tiered evaluation and testing effort is recommended: 1) performance modeling for prediction and analysis, 2) laboratory and simulator empirical studies for validation of both models and products, and 3) airborne flight tests to cross-validate simulator results.

References

Aume, Nils M. and Robert G. Mills (1981), The Results of AFAMRL Remotely Piloted Vehicle (RPV) Simulation Studies VII and VIII, AFAMRL-TR-80-98, Air Force Aerospace Medical Research Laboratory, Wright-Patterson Air Force Base, OH.

Bachert, Robert F., Kenneth H. Evers, Constance M. Hoyland, and Evan P. Rolek (1983), "Static and Dynamic Model Integration for a SAM C3 Simulation via IDEF0 / SAINT," Proceedings of the IEEE International Conference on Cybernetics and Society, Seattle, WA, 144-148.

Bent, N. E. (and others) (1991), The TAC BRAWLER Air Combat Simulation Analyst Manual (Rev. 6.1), Decision Science Applications, Inc., Arlington, VA.

Card, Stuart K., Thomas P. Moran, and Allen Newell (1983), The Psychology of Human-Computer Interaction, Lawrence Erlbaum Associates, Inc., Hillsdale, NJ.

Chubb, Gerald P. (1995a), Memorandum: Mission Area Plan Reviews, unpublished manuscript, AL/CFHI. Wright-Patterson AFB, OH.

Chubb, Gerald P. (1995b), "Displays to Enhance Situational Awareness," International Conference on Experimental Analysis and Measurement of Situation Awareness, Daytona Beach, FL, 1-3 Nov. 1995 (in press).

Chubb, Gerald P. (1992), Sequential Network Modeling (SNM) for Performance Risk Assessment: I. General Considerations, (draft monograph), US Army Office of Military Performance Assessment Technology (OMPAT), Silver Spring, MD, 30 September.

Chubb, Gerald P. (1991), "Proposed Methodology for Synthetic Task Construction," AFOSR SFRP Final Report 1991, Air Force Office of Scientific Research, Bolling AFB, MD.

Chubb, Gerald P. (1978), SAINT: A Combined Simulation Language for Modeling Large, Complex Systems, AMRL-TR-78-48, Aerospace Medical Research Laboratory, Wright-Patterson AFB, OH.

Chubb, Gerald P. and Richard S. Jensen (1994), Two Paradigms for Aeronautical Decision Making, DOT/FAA/AM-94/XX, Federal Aviation Administration, Aeromedical Division, Washington, DC.

Chubb, Gerald P. and Constance M. Hoyland (1989), "Systematic Modeling of Multioperator Systems," in McMillan, Beevis, Salas, Strub, Sutton, and Van Breda, editors, Applications of Human Performance Models to System Design, Plenum Press, New York, NY, 295-311.

Duket, Steven D., David B. Wortman, and Deborah J. Seifert (1976), SAINT Simulation of a Remotely Piloted Vehicle Drone Control Facility, AMRL-TR-75-119, Aerospace Medical Research Laboratory, Wright-Patterson AFB, OH.

Durlach, Nathaniel I. and Anne S. Mavor, Editors (1995), Virtual Reality: Scientific and Technological Challenges, National Academy Press, Washington, DC.

Grieme, R., David Cleveland, and Gerald P. Chubb (1969), "Project PIMO Final Report. Vol. IV: PIMO Technical Data Specification," AFHRL-TR-69-155, Air Force Human Resources Laboratory, Wright-Patterson AFB, OH.

Jensen, Richard S., J. Chubb, L. Kirkbride, and J. Adrion (1994), A Review of Human Judgment Literature with Aviation Implications, DOT/FAA/AM-94/XX, Federal Aviation Administration, Aeromedical Division, Washington, DC.

Johnson-Laird, P. N. (1983), Mental Models: Towards a Cognitive Science of Language, Inference, and Consciousness, Cambridge University Press, New York, NY.

Laird, John, Paul Rosenbloom, and Allen Newell (1986), Universal Subgoaling and Chunking: The Automatic Generation and Learning of Goal Hierarchies, Kluwer Academic Publishers, Boston, MA.

McNeese, Michael D., Brian S. Zaff, Maryalice Citera, Clifford E. Brown, and Randall D. Whitaker (1955), "AKADAM: Eliciting User Knowledge to Support Participatory Ergonomics," International Journal of Industrial Ergonomics, Vol. 15, No. 5 (May), 349-363.

Miller, George A., Eugene Galanter, and Karl H. Pribram (1960), Plans and the Structure of Behavior, Henry Holt & Co., New York, NY.

Newell, Allen (1990), Unified Theories of Cognition, Harvard University Press, Cambridge, MA.

O'Neil, Richard (1993), Education / Training Instructional System Development, AF Manual 36-2234, HQ AETC / XORR, Randolph AFB, TX.

Pritsker, A. Alan B. (1995), Introduction to Simulation and SLAM II, Fourth Edition, John Wiley & Sons, New York, NY.

Sanderson, Penelope M. (1994), MacSHAPA Version 1.0.2 Manual, Crewstation Ergonomics Research Information Analysis Center, University of Dayton Research Institute, Dayton, OH.

Sarter, Nadine B. and David D. Woods (1993), "Pilot Interaction with Cockpit Automation: Operational Experience with the Flight Management System," International Journal of Aviation Psychology, Vol. 3, No. 4, 303 - 321.

Sathe, Vijay (1985), Culture and Other Corporate Realities: Text, Cases, and Readings on Organizational Entry, Establishment and Change, Irwin, Homewood, IL.

Seifert, Deborah J. and Gerald P. Chubb. (1973), Computer Models of Man-Machine Survivability/ Vulnerability, AMRL-TR-72-69, Aerospace Medical Research Laboratory, Wright-Patterson AFB, OH.

Smith, S. L., B. Purvis, and S. Turner (1994), "Test Planning, Analysis and Evaluation System (TEST PAES): A Process and Tool to Evaluate Cockpit Design During Flight Test," Proceedings of the Human Factors Society 38th Annual Meeting, Nashville, TN, 942-948.

Warner, Norman, J. Jay Wolf, Stan Olejniczak, Estrella Forster, Michael Messick, Charlie Barba, and Jeffrey G. Morrison (1995), First Application of the Performance Metrics Methodology for Predicting Impact of Subsystem Technology on Aircraft Mission Performance, NAWCADWAR-95010-4.6, Naval Air Warfare Center, Aircraft Division, Warminster, PA.

Wiener, Earl L. (1988), "Cockpit Automation," in E. L. Wiener and D, C, Nagel, editors, Human Factors in Aviation, Academic Press, New York, NY, 433-461.

Wilmot, H. Leslie, Gerald P. Chubb, and Barbara Tabachnick (1960), "Project PIMO Final Report. Vol. VI: Technical Data Preparation Guidelines," AFHRL-TR-69-155, Air Force Human Resources Laboratory, Wright-Patterson AFB, OH.

Zaff, Brian S., Michael D. McNeese, and Daniel E. Snyder (1993), "Capturing Multiple Perspectives: A User-Centered Approach to Knowledge and Design Acquisition," Knowledge Acquisition, 5, 79-116.

Michael Dooley report not available at time of publication.

Daniel Ewert report not available at time of publication.

COMPARATIVE EFFECTS OF DYNAMIC AND STATIC STRENGTH TRAINING ON $+G_Z$ TOLERANCE

Ellen Glickman-Weiss
Associate Professor
Department of Kinesiology

Louisiana State University
121 Long Field House
Baton Rouge, LA 70803-7101

Final Report for:
Summer Faculty Research Program
Armstrong Laboratory

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, DC

and

Armstrong Laboratory

July 1995

COMPARATIVE EFFECTS OF DYNAMIC AND STATIC STRENGTH TRAINING ON $+G_Z$ TOLERANCE

Ellen Glickman-Weiss
Associate Professor
Department of Kinesiology
Louisiana State University

Abstract

The comparative effectiveness of generalized dynamic resistance strength training and specific static resistance strength training in enhancing relaxed, gradual onset rate (GOR) $+G_Z$ tolerance and $+4.5$ to $+ 7.0$ G_Z simulated air combat maneuver endurance is the focus of the present investigation. Because of the muscular efforts demanded of individuals during repeated performance of the AGSM during high +Gz conditions, physically untrained individuals will fatigue earlier than their physically trained counterparts. Thus, the protective effect of general strength training programs on $+G_Z$ endurance has indicated enhancement of $+G_Z$ endurance in strength-trained subjects. Additionally, results by the Russians indicated that static force generated by a pilot during a progressive test of sustained, static leg press was highly predictive of $+G_Z$ tolerance. Thus, it was the purpose of the present investigation to compare these training regimes as to serve as an important step in responding to the flying community's request for a physical training program which enhances $+G_Z$ endurance and also reduces aircrew training time involvement. The present investigation has not been completed, and is ongoing, thus no final results are reported at present.

COMPARATIVE EFFECTS OF DYNAMIC AND STATIC STRENGTH TRAINING ON +Gz TOLERANCE

Ellen Glickman-Weiss

Introduction

G-tolerance is the ability to maintain consciousness, and effective performance when under G-stress. In doing so, blood pressure and blood flow must be maintained. Factor's which may effect the individual's G tolerance include: 1) the effectiveness of the anti-G-straining maneuver, 2) physical, physiological and psychological factors, 3) equipment and, 4) G- discipline and knowledge. (AFPAM 11-404, p.4).

The anti-G straining maneuver (AGSM) has two principal components (Gillingham 1988). The first component is a muscular component, which requires the sustained, static recruitment of the skeletal musculature. Second, the intra-thoracic pressure component of the AGSM involves repeated 3-s cycles of forceful expiratory efforts against a closed glottis (L-1 maneuver, Cote et al. 1986, Gillingham 1988). The L-1 AGSM is the best G-defense available for aircrew members. Equipment was never meant to replace the AGSM, only aid it. Therefore, the repetition of an effective AGSM is critical to fighter aircrew in the sustained high $+G_z$ environment of air combat maneuvering.

Physical fitness programs in general, have been shown to enhance the individual's G-tolerance (Bulbulian et al. 1994, Burton 1986, 1987, Tesch et al. 1983) perhaps by enhancing the active muscles involved in the AGSM. More specifically, research has demonstrated weight training (i.e., anaerobic training) increases the individual's ability to withstand high G-forces for a longer duration with less fatigue. In addition, aerobic training has been shown to decrease the recovery time between centrifuge training runs. The present investigation appears to be critical to better understand the value of a specific strength training program on the individual's ability to withstand G-stress.

Methodology

Equipment and Facilities. Subjects' relaxed $+G_Z$ tolerance and +4.5 to +7.0 G_Z SACM endurance will be assessed from standards measurement equipment during exposures on the Armstrong Laboratory Centrifuge. Subjects will be experienced members of the acceleration subject panel and will be fitted with the standard CSU-13 B/P anti-G suit, HGU-55/P helmet and MBU-12/P oxygen mask for all centrifuge runs. To minimize breathing resistance from the oxygen mask during the AGSM, a high-flow CRU-93 oxygen regulator, in the non-pressure breathing setting, will be used for all centrifuge runs.

All subjects will be fitted with a 5-lead ECG for two-axis recording of heart rhythms and heart rate during all centrifuge exposures. Analog data signals of physiologic and anti-G suit pressure variables will be recorded on a multiple channel strip recorder (Gould 2800S), with magnetic tape back-up, for subsequent data reduction and statistical analysis. Centrifuge subjects will be observed via closed circuit video monitors in the centrifuge control room.

An ACES II upright seat (13° seatback angle), such as used in the F-15 aircraft, has been modified for use as the static strength training device. This "static trainer" is equipped with individual foot boards fitted with strain gauge transducers which interface with a microprocessor, allowing continuous measurements of static leg press efforts (displayed in kg) during static strength testing and strength training sessions.

With the exception of biceps curl exercises performed with barbells, subjects assigned to the dynamic, strength training program will be trained and tested on a multi-station exercise apparatus (Universal Gym). A noninvasive arterial blood pressure monitoring system (Finapres) also will be used for all subjects during tests of static and dynamic leg press strength.

A cycle ergometer (Monark 818E) will be used for subject performance of tests of aerobic capacity (VO_2 peak). Peak anaerobic power and mean anerobic power will be determined from administration of a 30-s supramaximal pedaling test on a different cycle ergometer (Cardionics, Sweden). VO_2 peak and respiratory exchange ratio (RER) values will be determined from analysis

of expired gas (Sensormedics 2900z Metabolic cart) collected from subjects during tests of anaerobic capacity and anaerobic power. Cycle ergometry tests will be conducted in the Exercise Physiology Laboratory in Building 110.

A pressure manometer (Boehringer) with negative inspiratory force monitor and pressure gauge adapter (Instrumentation Industries, Inc.) will be used to measure maximal expiratory pressure and maximal (negative) inspiratory pressures of all subjects. These tests will be performed by pulmonary function technicians in Building 110.

Subjects. Male and female subjects from the Armstrong Laboratory acceleration subject panel who have maintained acceleration training standards will be recruited for this study. We will recruit sufficient subjects to have approximately 14 subjects complete the study. Subjects must have recent $+G_Z$ exposure experience and be able to complete at least four cycles (135-s) of the +4.5 to +7.0 G_Z SACM to qualify for the study. Because acceleration panel subjects have not routinely performed SACM exposures with helmet and mask, subjects in the study will perform SACM familiarization runs wearing the full equipment ensemble, including helmet and mask, prior to establishment of maximal baseline SACM endurance times. Panel subjects will also receive exercise program indoctrination before they perform the strength testing and initiate the respective 8-week training regimens. Only acceleration panel subjects who have no history of prolonged strength training participation and have not participated in muscular strength training programs during the past four weeks will be studied. To control for the potentially detrimental effects of excessive aerobic training on $+G_Z$ tolerance (Bulbilian et al. 1994), subjects who perform regular aerobic activities equal or equivalent to running more than 15 miles per week also will not be included in the study.

Duration of the study. Determination of relaxed $+G_Z$ tolerance, maximal SACM endurance, static or dynamic strength (with Finapres blood pressures), aerobic capacity, anaerobic power, and maximal inspiratory and expiratory pressures will require approximately one week. These determinations will be performed prior to (pre) following (post) each 8 -week strength training regimen. Subjects will participate in both strength training regimen (cross-over

experimental design), with a 4-week detraining period between regimens, beginning after completion of post training testing for the first 8-week period. Duration of participation for each individual subject will be approximately 22 weeks.

General design and Overview. The general experimental design is a 2 x 2 repeated measures factorial design, blocked on subjects. There are two levels of treatment (general/dynamic & specific/static) and two levels of tests (pre & post) for measurement of physiologic variables accomplished prior to and following each 8-week training regimen. Analysis of variance (ANOVA) and tests of pairwise comparisons will be used to test for significant change in mean scores of dependent variables over the 8-week training regimens. Dependent variables include:

a. relaxed gradual (0.1 G/s) onset rate (GOR) $+G_Z$ tolerance;

b. peak heart rate response to +4.5 to +7.0 G_Z SACM (6 G/s)

c. SACM endurance (s)

d. subjects reported perceived effort (scale 1-10) during SACM endurance runs

e. aerobic capacity (VO_2 peak)

f. peak anaerobic power and mean anaerobic power.

g. static muscular strength from the static leg press test

h. dynamic muscular strength (1 RM; in kg) and;

i. systolic and diastolic blood pressures obtained with Finapres during both static and dynamic tests of leg press strength.

In order to maintain their entry AGSM proficiency and $+G_Z$ endurance baseline, subjects will be required to perform a +4.5 to +7.0 SACM centrifuge training profile at least once but not more than twice weekly. This $+G_Z$ training requirement consists of a rapid onset rate (ROR) 5 G/10-s warm-up, followed by four cycles (@135-s) of the SACM. Pre-and post-regimen tests of maximal SACM endurance will satisfy the weekly SACM proficiency requirements for those particular weeks. Weekly AGSM proficiency centrifuge exposures will continue during the respiratory muscular training regimen.

Day 1 of testing/subject orientation involves establishment of relaxed $+G_Z$ tolerance and +4.5 to +7.0 G_Z SACM endurance on the centrifuge. On day 2, subjects will be tested for aerobic capacity (VO_2) peak on a cycle ergometer. Following an orientation to the strength training programs on day 3 subject will be tested on day 4, for anaerobic power using an electrically braked cycle ergometer (Bar-Or 1978) . Finally, on the morning of day 5, maximal expiratory and inspiratory pressures will be tested in the pulmonary function facilities of Building 110. These tests will establish subject inspiratory and expiratory pressure baselines and be used to indirectly assess changes in respiratory muscle strength after completion of the 8-week strength training periods. Following day 5 pulmonary testing, subjects will perform a final rehearsal of the strength training regimen. At least two but not more than seven days following the various testing and strength training orientation sessions, subjects will initiate either of the two 8-week muscular strength training programs. Subjects tests and orientation will be performed in the order described.

Subjects next will be assigned in a balanced, randomized order, to either the dynamic strength training regimens or the static strength training regimen. An exercise supervisor will indoctrinate all subjects on technique training for their particular exercise regimen prior to testing subject performance on regimen-specific strength tests. With the exception of the use of dunbbells for testing (and training) of biceps strength and a 1-minute "abdominal crunch" test to assess abdominal strength, dynamic strength testing will consist of determining each subject's one repetition maximum (1RM) lifting capability for each lift on the strength training apparatus. In addition to the abdominal and neck exercises previously described, subjects in the dynamic exercise program will perform a typical body weight-training exercise on a Universal Gym apparatus. Specific exercises are as follows:

leg press	supine bench press
knee extension	arm (triceps) push-down
knee flexion (leg curl)	lat (latissmus) pull-down

Subjects static strength will be determined from the best of three maximal static leg press efforts (5-s duration) performed in the static trainer. As the second measure of static strength (static endurance), subjects will be tested on the duration (0-150 s) they can sustain the progressive 5-level static strength training task in the static trainer.

Dynamic and static strength sessions will be conducted three days per week,on alternate days. Dynamic strength training will consist of each subject's performing to fatigue on weight loads established during the first days indoctrination session. Weightloads for each exercise will be calculated as 80% of the individual's 1RM, a weight which theoretically can be lifted no more than 5-7 repetitions (5-7RM). For each static strength training session, subjects will perform a single 150-s "set" (5 x 30-s phases) of a sustained, progressively more difficult static leg press in the static training device interfaced to a computer display.

Within each 8-week training regimen, if a subject fails to participate in all three of the supervised exercise sessions per week on more than one occasion, or if the individual fails to participate in less than two sessions per week on any occasion the subject will be disenrolled from the study. After completing the first 8-week exercise training regimen , all subjects will be tested for changes in $+G_Z$ tolerance and associated physiological data, maximal inspiratory and expiratory pressures, peak VO_2, anaerobic power, and dynamic respiratory muscular strength. Following all post-regimen testing subjects de-train for four weeks. Subjects will, however, continue their 4-cycle SACM centrifuge exposure. After the 4-week period of de-training, all physiologic and physical tests will be collected and the subject will be indoctrinated on the new (crossover design) strength training regimen before the subjects initiate their final 8-week strength training regimen.

Centrifuge Procedures. Although peak relaxed $+G_Z$ tolerance will be established and periodically re-assessed for all subjects, the principal criterion by which effectiveness of the exercise conditioning regimens will be evaluated is the duration, in seconds(s), subjects can endure the +4.5 to +7.0 G_Z SACM profiles. For training and data collection runs, the centrifuge will be configured with the upright (13° seatback angle) ACES II-type seat. Foot plates in the centrifuge gondola will be positioned to provide an approximately 120° knee joint angle. Each subject will be properly fitted

with a standard CSU-13 B/P anti-G suit and a standard HGU-55/P helmet with MBU-20/P mask for all centrifuge exposures. Acceleration profiles for pre-and post-experimental data collection consist of:

a. GOR (1 $+G_Z$/s) to relaxed peak $+G_Z$ with standard anti-G suit on but uninflated;

b. One cycle of ROR (6 $+G_Z$/s) to +5 G_Z for 15-s (warm-up), followed by 2-min rest;

c. standard +4.5 to +7.0 G_Z SACM to fatigue or other end-point criteria [medical monitor decision, self-initiated decision, nausea, 100% peripheral light loss (PLL), 50% central light loss (CLL), or G-induced loss of consciousness (G-LOC)];

d. 5-min resting recovery in gondola;

e. repeat SACM profile, step (c); times for both trials will be analyzed separately.

Testing of Aerobic Capacity and Anaerobic Power. Each subject will perform cycle ergometry exercise tests to determine both aerobic capacity and anaerobic power output (Bar-Or 1978) . For both tests, subjects will report to the Exercise Physiology Laboratory (Building 170) dressed in clothing suitable for exercise.

The subject's VO_2 peak (peak aerobic capacity) will be determined while the individual performs an incremental cycling protocol on a cycle ergometer. Every two minutes, the workload will be increased until the subject reaches volitional fatigue. Achievement of VO_2 peak will be determined when at least two of the following criteria are met: leveling of oxygen consumption (VO_2); respiratory exchange ratio (RER) above 1.15; attainment of predicted maximum heart rate. Gas exchange will be assessed using the Sensormedics 2900z Metabolic Cart for determination of VO_2 and RER values.

The Wingate Anaerobic Test is a 30-s, maximal effort, cycle ergometer task (Bar-Or 1978).

The warm-up for the test consists of 3-5 min of pedaling at a moderate load (@175 Watts), interspersed with two sprints of 3-5 s each. For the test, the subject pedals as fast as possible against a workload (.075 kg x kg body weight $^{-1}$) which requires a maximal effort for 30-s. Recovery includes 1 min of pedaling against no resistance, followed by 4 min of pedaling against a light-to-moderate resistance (approximately 100 Watts).

Both the VO_2 peak (aerobic) and Wingate (anaerobic) tests may be terminated because: 1. the subject chooses to stop the test for any reason; 2. the subject reaches volitional fatigue; 3. indications of abnormal cardiovascular response are present; 4. the subject experiences light-headedness, pallor, nausea; or 5. the equipment fails.

Static Strength Training and Testing Procedures. The static strength training regimen entails supervised performance of progressively increasing static muscular contractions of the hip and knee extensors (leg press) against fixed foot boards in the modified ACES II seat ("static trainer"). After adjusting the seat to establish a 120° knee joint angle, each subject will perform a general calisthenic warm-up before performing the initial 5-step (150 s) static leg press exercise set. Preparatory to initiating the training regimen, the subject will secure himself herself in the static trainer seat and perform a single 5-s static leg press effort (specific warm-up) at 120 kg, as presented on computer-generated visual prompts. At one minute after completion of the warm-up, the subject will continue to follow computer prompts to initiate and maintain the progressive static leg press exercise training regimen. Subjects will press simultaneously on left and right foot pedals fitted with strain gauge transducers which transmit force signals to a microcomputer. Processed signals then are converted into kilogram (kg) force signals and recorded for subsequent reduction and data analysis. A computer-generated video bar display will indicate the static force the subject produces. The subject's registered effort must not vary by more than ±5% from the required static effort displayed on the video. The successive, continuous progression of 30-s static efforts (120, 160, 200, 240, and 280 kg) continues until the subject completes 150 s of static leg press effort.

During an initial static strength test, the greatest of the forces generated during three maximal leg press force efforts will be recorded as the static strength value for the leg press. Maximal static efforts will not exceed 5.0 s, and a 2-min resting recovery will be provided between efforts. An additional test of static strength/endurance requires subjects to perform their best static leg press regimen, performed according to computer-generated prompts. As in training, during this strength test, the subject will attempt to maintain force levels of 120, 160, 200, 240, and 280 kg for 30 s at each of the five respective prescribed force levels. The test is completed when the subject completes the entire 150s or (more likely) is unable to maintain the solicited force requirement . A subject's score will be the number of seconds (0-150) he/she is able to sustain the force displayed on the feedback monitor. A standard 10-minute recovery will be required between the maximal leg press strength test and the 150-s sustained force test to allow subjects to recover from the previous test.

General, Dynamic Strength Training and Testing Procedures. Training and testing of abdominal strength/ endurance will be accomplished by performance of trunk flexion exercises ("abdominal crunches"), according to USAF published guidelines (AFP 11-404). Abdominal strength / endurance training consists of three sets of abdominal crunches to fatigue. Self-resisted neck strength exercises ("towel exercises") will be performed by subjects in both exercise groups to reduce the risk of neck injury during centrifuge exposures. However, because of the added risk of neck injury associated with maximal testing, neck strength will not be tested. Except for abdominal crunches, neck exercises, anddumbbell use in training and testing elbow flexion (biceps curl) strength, dynamic strength training and testing will be performed on a multi-station weight stack apparatus (universal Gym). A 6 to 8 repetition warm-up at relatively light weightloads (@50% of 1 RM) for each lifting exercise will be performed by all subjects prior to the first exercise set during training and before initiating heavy lifting during strength tests to determine the 1RM. Exercises consist of leg press, knee extension, knee flexion (leg curl), supine bench press, arm (triceps) pushdown, and lat pull-downs.

Abdominal strength/endurance will be determined as the number of abdominal crunches properly performed in a 60-s set. Arm flexion (biceps) strength will be assessed from 1 RM tests with a barbell (free weights). Other dynamic strength testing consists of determining the 1RM weight load for each of the dynamic exercises performed on the weight stack apparatus. All tests will be preceded by a light resistance warm-up of 6 to 8 repetitions. Next, starting at a weight load the exercise supervisor estimates as approximately 75% of the subject's 1 RM, the subject will attempt single lifts at several progressively greater weight loads until a weight is attempted he/she is unable to lift. After each accomplished lift, the supervisor will increase the weight as the subject rests for 60 s. As the weight load approaches the subject's maximal lift capability, the supervisor will make smaller weight load increases and the subject's rest period will increase to 90. When a weight is attempted which the subject is unable to lie, he/she will rest for 90 s before making a second attempt at the same failure weightload. The highest weight successfully lifted is recorded as the subject's 1RM score. The 1 RM should be determined by not more than the fifth or sixth weight increase. This procedure will be repeated for all dynamic exercises. Strength tests conducted at the end of the 6-week training regimen will be initiated at approximately 1.15 times the subject's most recent 5-7RM training weight load.

Dynamic strength training will be conducted in three sets at 80% of the 1RM (@ 5 to 7RM) weight load. In order to maintain the progressive resistance format of strength training, supervisors will add weight, as appropriate, to the training weight load once a subject demonstrates his/her ability to perform more than eight repetitions (8RM) on the third set of that exercise. Additionally, supervisors will encourage subjects to perform to fatigue on every set of all exercises. A rest interval of 60 seconds will be provided between sets of all dynamic liking exercises. Exercise supervisors will encourage subject compliance and will record subject attendance and performance (RM workload and repetitions) for all exercises.

Statistical Analysis: The 2 x 2 factorial design incorporates an analysis on variance (ANOVA) with pairwise comparisons to test for significant differences between mean scores of the dependent variables according to the two levels of strength training regimens and two levels (pre and post) of test time. Regression analysis also will be performed to determine if $+G_Z$ SACM endurance times can be predicted from strength scores obtained from subjects prior to each strength training regimen. Additionally, correlation tests (Pearson) will be run to determine if significant relationships exist between means or changes in means of pertinent dependent variables.

Results, Conclusion

The present investigation is currently in progress thus, the results and thereby the conclusions can not be reported. The duration of this investigation is 22 weeks, thus the project and its results will be forthcoming.

References

Air Force Pamphlet 11-404, Flying Operations. G-Awareness for Aircrew, 19 August 1994 Section C-Application.

Bar-Or O. A new anaerobic capacity test-characteristic and applications. Proceedings of the 21st World Congress in Sports Medicine. Sep 7-12, 1978, Brasilia.

Bulbulian R, Crisman RP, Thomas ML, Meyer LG. The effects of strength training and centrifuge exposure on $+G_Z$ tolerance. Aviat Space Environ. Med. 1994; 65: 1097-1104.

Burton RR. G-induced loss of consciousness: definition history, current status. Aviat Space Environ. Med. 1987; 59; 2-5.

Burton RR. Simulated aerial combat maneuvering tolerance and physical conditioning: current status. Aviat Space Environ. Med. 1986; 57; 712-4

Cote R, Tripp L, Jennings T, Karl A, Goodyear C, Wiley R. Effect of inspiratory volume on intrathoracic pressure generated by an L-1 maneuver. Aviat. Space Environ. Med. 1986; 57: 1035-8.

Gillingham KK. High G stress and oriental stress: physiological effects of aerial maneuvering. Aviat. Space Environ. Med. 1988; 59 (suppl) A10-A20.

Lake FR, Henderson K, Briffa T, Openshaw J, and Musk AW. Upper-limb and lower limb exercise training in patients with chronic airflow obstruction. Chest 97: 1077-1082, 1990.

Tesch PA, Hjort H, Balldin UI. Effects of strength training on G tolerance. Aviat Space Environ. Med. 1983; 54(8) 691-695.

Development of a Fluorescence Post-Labeling Assy for DNA Adducts

Joseph B. Guttenplan
Professor
Department of Basic Sciences

New York University
New York, NY

Final Report for:
Summer Research Program
Armstrong Laboratory

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, Washington DC

and

Armstrong Laboratory

September 1995

Development of a Fluorescence Post-Labeling Assay for DNA Adducts

Joseph Guttenplan
Departments of Biochemistry and Environmental Medicine
New York University Dental and Medical Schools
New York, NY 10010

Abstract

Methods for detecting low levels of relatively small DNA adducts are currently limited. Probably the most frequently used method is a ^{32}P-post labeling assay. This assay, although very sensitive for large hydrophic DNA adducts, is less sensitive for smaller adducts and has the disadvantage that it requires the use of considerable radioactivity. Adducts likely to be formed by compounds of interest to the Air Force will probably fall into the class of smaller adducts. Thus, there is a need for an alternate assay. One potentially sensitive method for detecting such adducts is a fluorescence post-labeling assay. This method would couple a highly fluorescent label to deoxynucleotides (dNMPs) prepared by digesting the DNA. The fluorescent nucleotides would then be analyzed by HPLC. Because the label can be selected for a high quantum yield of fluorescence, the method should be extremely sensitive (low fmol detection). When the project was initiated there was only one reported method for fluorescent labeling of dNMPs and this was the starting point for the project. That method utilized a carbodiimide coupling with dNMPs, the replacement of the carbodiimide with ethylenediamine (ED) and the derivitization of the free amino group with dansyl chloride. The method was carried out using normal dNMPs, but proved less than satisfactory. The major difficulty with the method was interference from the hydrolysis product of dansyl chloride (dansyl sulfonic acid) and from the reaction product of ED with dansyl chloride. In order to reduce the concentrations of these products, the excess ED was removed. This was accomplished by releasing the free base and lyophilizing the mixture before addition of the dansyl chloride. This not only reduced the concentration of the ED-dansyl chloride product, but also allowed a dramatic reduction in the concentration of dansyl chloride. This was possible because the concentration of this reagent had to be in excess of the concentration of free amino groups, and the bulk of these came from the excess ED. A further improvement in the assay was brought about by utilizing a different fluorescent label, Fluoram. In contrast to dansyl chloride, this compound yields a nonfluorescent hydrolysis product. As a result the major interfering peak was eliminated and a very clean chromatograph was obtained.

In order to test the modified method it was necessary to synthesize appropriate standards of modified dNMPs. These were produced by reacting high concentrations of either chloroacetaldehyde (CA) or chloral hydrate (CH) with the dNMPs. CA is known to from DNA adducts, but less is known about CH. HPLC analysis of the reaction products revealed that CA and to a lesser extent, CH gave rise to modified dNMPs in sufficient yield to be collected for use as standards. Thus, at this point the methodology has been worked out to evaluate the fluorescence post-labeling assay on DNA adducts of the type produced by chlorinated aliphatic compounds.

DEVELOPMENT OF A FLUORESCENCE POST-LABELING ASSAY FOR DNA ADDUCTS

Joseph Guttenplan

Introduction

Carcinogenesis is believed to be a multistep process commencing with modification of, or damage to DNA, followed by the conversion of some of this damage to mutations and other permanent alterations in DNA (1,2). When mutations occur in certain positions in protooncogenes and/or tumor suppressor genes they may lead to tumor formation after promotion and progression (3,4). Therefore, the presence carcinogen-modified DNA in cellular DNA represents one risk factor for carcinogenesis, and agents capable of producing modifications in DNA represent potential carcinogens.

There are a number of methods for detecting DNA modifications (5). For the purpose of biomonitoring in humans the ^{32}-post-labelling is the more frequently used method (5). In carcinogen evaluation in experimental animals, and in <u>in vitro</u> systems the administration of radiolabelled carcinogens followed by detection of radioactivity in the DNA or DNA digests is frequently employed as well as the post-labelling assay. When carcinogens are unavailable with a radioisotopic label, or the expense precludes whole animal exposure, HPLC with uv, fluorescent or electrochemical detection is often employed. Detection of fluorescent DNA adducts can be extremely sensitive, but most DNA adducts do not fluoresce with sufficiently high quantum yields to allow this method to be generally applied.

The ^{32}-post-labeling assay is generally the most sensitive method for the detection of DNA adducts, with detection levels often reported in the range of one modified nucleotide per 10^{6}-10^{-9} nucleotides (5). Even lower levels of adducts are claimed to be detectable, but reports of such low levels are infrequent, probably because of technical difficulties encountered at such high sensitivities. The assay is much more sensitive for very hydrophobic adducts than smaller less hydrophobic adducts, and indeed most reports employing this method have focussed on polycyclic aromatic hydrocarbon or amine adducts (5). A major disadvantage of the assay is the requirement for relatively large quantities of radioactivity. Another is that adaptation of the method for smaller adducts usually requires a standard of the expected adduct, and the conditions of the assay must be tailored to the individual adduct. Thus, the method is not well suited as a general method for detecting unknown smaller DNA adducts. A potential improvement in assays for less hydrophobic adducts would be a general fluorescent post-labelling assay. In such an assay the adduct itself need not be fluorescent, but would be derivatized with a highly fluorescent group. Fluorescent derivatizing agents are commercially available and exhibit very high quantum yields of fluorescence. With such derivatives sub-picogram quantities adducts would be detectable by HPLC, resulting in sensitivities approaching those of the ^{32}P-post-labelling assay. However, the available fluorescent reagents do not react with DNA or its digestion products. Thus far, there is only one laboratory which has reported the development of a method to couple these reagents to nucleotides (6,7), but it difficult to determine whether that

method is applicable for the DNA adducts likely to be formed by compounds of interest to the Air Force, because certain regions of published chromatographs are obscured by fluorescent byproducts.

The goal of this Air Force Summer Faculty Research Project was to develop an improved fluorescent post-labelling assay suitable for the analysis for the types of adducts possibly formed by compounds of interest to the Air Force. To accomplish this goal the one reported method would have to be improved and standards of adducts similar to potential adducts of interest would have to be synthesized, and analyzed by the new method.

The fluorescence post-labeling assay could then be used to analyze DNA from organs of animals treated with potential carcinogens, or from <u>in vitro</u> systems (e.g., hepatocytes, tissue slices, etc.) treated with carcinogens. It could also be used in predictive toxicology since, in principle, it could detect many adducts formed by carcinogens with DNA <u>in vitro</u>, and the results from the analyses with a series of related structural could then be used as the basis for structure-activity relationships which could then be applied to new or untested compounds.

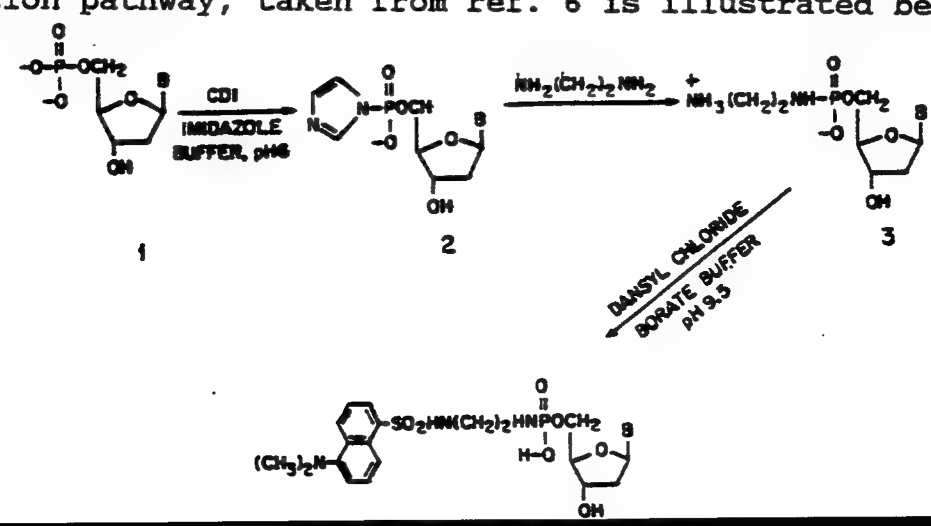
Methods

Equipment

HPLC was performed on a system from ESA Inc. (Chelmsford, MA). The system consisted of a Shimatzu pump, a Triathelon autoinjector, a Linear Instruments variable wavelength uv detector and a Linear dual-monochrometer fluorescence detector. For some injections a Rheodyne manual injector was employed. Data collection and storage was handled by a Digital Instruments computer. The initial period of the project was spent assembling and learning to operate the system.

Formation of phosphoramidates

The fluorescent post-labelling method as described in the literature involves digestion of DNA to deoxynucleotides (either 3′ or 5′monophosphates) followed by reaction at pH 6.0 with the water soluble carbodimiide, 1-(3,3-dimethylaminopropyl)-3-carbodiimide generating a phosphorimidazolidate. On reaction with an amine the carbodiimide is displaced resulting in a phosphoramidate. If the amino substituent contains another nucleophilic group (in this case a second amino group) this second group can serve as a site for coupling with an electrophilic fluorescent reagent. The common fluorescent reagent, dansyl chloride was previously utilized (6,7). The reaction pathway, taken from ref. 6 is illustrated below.



In order to develop this method for the Air Force project, the initial formation of the phosphoramidates of each of the four 3'dNMPs was first investigated. As described in the literature the product was formed by incubating the dNMP (0.1-2mM), CDI (0.15M) and ethylenediamine (0.20M) in 0.1M imidazole buffer for eight hours at room temperature (6). This first two steps of the reaction were carried out without isolating the intermediate product (6,7), and reactants and products were analyzed by reverse phase HPLC in 50mM phosphate (pH 6.0) containing 0-10% methanol (depending on the dNMP). Fig. 1a shows the 3'dCMP reaction mix about 15 min after mixing and Fig 1b shows the reaction after eight hr at room temperature. Although the reference article indicates the reaction is complete after 6 hr, it was actually somewhat less (ca. 90%) than complete after 8 hr. Similar results were obtained with the other dNMP's (results not shown). It was felt that the yield was high enough to proceed with next step in the reaction.

Formation of fluorescent derivatives of the phosphoramidates

The phosphoramidates of the nucleotides were reacted with a 100mg/ml (in acetone) solution of dansyl chloride. Because the free amino group of the ethylenediamine must be unprotonated for efficient reaction with dansyl chloride, the pH was brought to 9.5 with 1M NaCO3 (pH 10.0). The concentration of dansyl chloride in the mix was brought to 0.2M and the mix was stirred for one hr at room temperature. A precipitate immediately formed (presumably dansyl chloride), and this was subsequently removed by passing a ten fold dilution of the reaction mix through a centrifuge filter.

In later experiments, the reaction mix was brought to a pH of 10.5 with $NaCO_3$ and lyophilized and redissolved in water before the dansyl chloride was added. As this procedure greatly reduces the large excess of ethylenediamine, it was not necessary to use such high concentrations of dansyl chloride and the volume of the dansyl chloride solution added was reduced 10 fold.

In addition to the dansyl chloride derivitization a second fluorescent derivative was prepared. The derivative of the lyophilized phosphoramidates was produced by reaction with fluorescamine (Fluoram). The lyophilized mix was redissolved in the initial volume of water and 1/4 volume of a 10mM solution fluorescamine in acetonitrile was added, the mixture rapidly mixed and then centrifuged. According to the manufacturer reactions between fluoram and primary amines are extremely rapid and do not require additional incubation.

Controls for both fluorescent reagents which contained all components except dNMP were carried through all steps in parallel. The reaction mixtures were analyzed by reverse-phase HPLC with fluorescence detection using a mobile phase of 50mM phosphate buffer (pH 6.0) containing 18-28 % acetonitrile (depending on the product).

Preparation of modified nucleotides

In order to test the fluorescence method for detection of DNA adducts, modified nucleotides of the types which can be formed between certain chlorinated aliphatic compounds and DNA were

prepared. These could subsequently be used in mixtures with normal nucleotides to evaluate the sensitivity of the fluorescence method. Two compounds were employed, chloracetal (chloroacetaldehyde-diethylacetal, which is in equilibrium with chloroacetaldehyde) and chloral hydrate. Solutions (0.3M) of these were both reacted with each of the four normal nucleotides (1-3mM). The chloracetal solutions were incubated for 2-8 hr at 45^0 C and the chloral hydrate solutions were incubated for 24 hr at 45^0. The mixtures were then analyzed by HPLC on a 4 x 250 mm C_{18} reverse phase column, using a mobile phase of 50mM phosphate buffer (pH 6.0) containing 0-10% methanol, depending on the dGMP.

Results

After formation of the phosphoramidates of the dNMPs and adjustment of the pH to 9.5, dansyl chloride was added to the mix, the reaction was allowed to proceed and the diluted, filtered solutions were analyzed by HPLC. The samples of the mixes containing each of the dNMPs were chromatographed and each showed two major peaks. The samples all showed two smaller peaks preceding the largest peak, and in addition, a third minor peak between the major peaks. This last minor peak exhibited a different retention time for each of the dNMPs, indicating it represents the dNMP-dansyl chloride derivative. As an example, the dCMP reaction products are shown in Fig. 2a. The first major peak is dansyl sulfonic acid, the hydrolysis product of dansyl chloride. The second major peak is the ethylenediamine-dansyl chloride reaction product (possibly an unresolved mixture of the mono and diamino adducts). The small peak between the major peaks is the dCMP-dansyl chloride derivative. Similar results were obtained with the other nucleotides (except the retention times of the dNMP derivatives were different (results not shown). Because the two major peaks tended to obscure large regions of the chromatograph and would have made detection of minute levels of DNA adducts difficult, an attempt was made to reduce the levels of the major peaks.

The reason for the high concentration of obscuring fluorescent products was the necessity to use sufficient dansyl chloride to derivatize all of the free amino groups. The bulk of these amino groups resulted from the large excess of ethylenediamine necessary for near-quantitive formation of the phosphoramidate. In order to selectively remove the excess ethylenediamine from the phosphoramidates of the dNMPs, advantage was taken of the fact that the free base of ethylenediamine is volatile. Therefore, before the derivatization with dansyl chloride the pH of the mix was raised to 10.5 and the mix was taken to dryness on a speed-vac. It was then possible to add considerably less dansyl chloride; and the ratio of the dCMP derivative to nonproductive fluorescent products was greatly increased (Fig. 2b). (Note: the noise in the chromatogram resulted from a crack in the flow cell).

In order to further reduce interference from fluorescent byproducts a different fluorescent reagent, Fluoram (fluorescamine) was tested. This reagent has the advantage that the product of the

excess reagent with water is not fluorescent. Fluoram was reacted with the pH-adjusted, lyophilized dNMP-phosphoramidate reaction mixtures and only two major peaks were observed. The reaction mix from TMP with Fluoram is shown in (Fig. 3a). The second peak was also present in a control in which the ED alone was reacted with fluoram (Fig. 3b) and was not present when ED was omitted from the reaction mix (results not shown). Therefore this peak presumably represents the ED-Fluoram product. Significant is the absence of any fluorescent product resulting from the Fluoram alone. The corresponding product from dansyl chloride represented the largest of the fluorescent peaks in Fig (2a and b). In an attempt to further reduce the ED background, the lyophilized phosphoramidate reaction mix was taken through another round of dissolution and drying. The residual mixture was redissolved and treated with Fluoram. The mix from the dCMP-phosphoramidate/Fluoram reaction is shown in Fig (4). This procedure resulted in an almost complete disappearance of the Fluoram-ED peak under the conditions employed here, and a single major peak corresponding to dCMP-Fluoram derivative was observed. Similar results were obtained using the other three dNMPs.

Preparation modified 3'dNMPs

In order to test the fluorescence-post-labelling method on modified dNMPs, a source of such compounds would be needed. Because chlorinated aliphatics represent one class of compounds of interest to the Air Force chloracetaldehyde, a compound reported to form DNA adducts (8) and also a metabolite of vinyl chloride (8), was employed. Each of the 3'-dNMPs was treated with a high concentration of chloroacetaldehyde and the products of the reactions were analyzed by HPLC using UV detection. Under the conditions employed dCMP and dAMP formed detectable products with chloroacetaldehyde (Fig. 5a,b and 6a,b). The yields were in the range of several percent. No corresponding peaks were detected when chloroacetaldehyde was incubated alone (results not shown). Both of these modified dNMPs were then suitable for standards for use in developing the fluorescence post-labelling assay. In addition, chloral hydrate, a pharmaceutical and a metabolite of trichloroethylene, was also tested as a possible source of DNA adducts. Chloral hydrate was also incubated with all four dNMPs, (but for about twice as long as with chloroacetaldehyde), and analyzed for products as described above for chloroacetaldehyde. A new peak was observed for the reaction mix containing dAMP (Fig. 7a,b). This product would then be suitable as a possible marker in identifying DNA adducts formed by chloral hydrate and trichloroethylene.

Discussion

The goal of the project was to develop a fluorescence post-labeling assay for the detection of DNA adducts of the type that could be formed by compounds of interest to the Air Force. There were three specific aims to the project: 1) to develop a

fluorescent post-labelling protocol, 2) to synthesize modified dNMP standards and 3) to determine the limits of detection of the fluorescently labelled modified dNMPs, in the presence of a large excess of normal dNMPs

As a starting point the one potential method already in the literature was taken, and that method was to be used to label normal 3'-dNMPs. The method needed modifications for the purposes required here. As a next step modified dNMPs were synthesized by reacting the normal dNMP's with compounds of interest to the Air Force. Only the last specific aim was not accomplished, due to a lack of time and some technical problems with the instrumentation.

The initial period in the laboratory was spent assembling and learning to use the HPLC system with fluorescence and UV absorbance detectors. The first experiments utilized conditions described in the original report from the literature. Both 3' and 5' dNMPs were reported to form phosphorimidazolidate derivatives which then reacted with ED to form phosphoramidates. For this project, the 3'dNMPs were used, as these are nucleotides which are labeled in the ^{32}P-post labeling assay. Therefore, a direct comparison of the sensitivities of the two assays should be possible. A study of the formation of the phosphoramidate reactions indicated that this product is formed from the dNMPs in high, but not quantitative yield. Although it should be possible to increase the yield, it was felt this would only lead to a small increase in the sensitivity of the assay and such "fine tuning" was at this stage, not necessary.

The phosphoramidates of the dNMPs all reacted with dansyl chloride to form fluorescent products, as previously reported (6). However, due to interference from the large fluorescent peak (dansyl sulfonic acid) resulting from the hydrolysis of the labeling reagent with water, and the large peak from the reaction of the excess derivatizing reagent, ED, with dansyl chloride, the method was less than satisfactory. The labeled nucleotides eluted on a shoulder of the dansyl sulfonic acid peak. With concentrations of normal dNMPs below about 10uM in the labeling reaction mix it became difficult to resolve their peaks. A much higher sensitivity would be required. For a typical DNA sample of around 100ug (ca. 3nmol dNMPs) which contains adducts at a modest level of 10^{-6} per nucleotide, the concentration of adducts in a 10ul reaction mix would be about 300pM; many orders of magnitude below the above limit of resolution. In addition, any adduct peaks which coeluted with dansyl sulfonic acid or the ED derivative of dansyl chloride would be totally obscured. Therefore, it was necessary to modify the published protocol.

Two modifications were introduced which dramatically increased the sensitivity of the assay. The first was the removal of the excess ED. The extra ED is necessary for the near-quantitive formation of the phosphoramidate of the dNMPs. However, since ED was present at about 200mM in the initial reaction mix and the adducts might be present at about 300pM it is apparent that a major reduction in the concentration of ED would be necessary. Because sufficient dansyl chloride must be added to derivatize all of the amino groups, reducing the concentration of ED would not only

reduce the concentration of the ED-dansyl chloride derivative, but also that of the unreacted dansyl chloride, since a slight fractional excess of the reagent is necessary to assure that the reaction is quantitative. After the phosphoramidate formation the excess ED serves no purpose and can therefore be eliminated. In order to remove this excess the ED was converted to its free base by raising the pH to 10.5 and then reducing the reaction mix to dryness. It was subsequently found that the concentration of residual ED could be further reduced by another round of dissolution and drying. With the concentration of ED greatly reduced, it was possible to reduce the concentration of dansyl chloride. This greatly reduced the interference from the byproduct peaks. Another significant improvement was obtained by changing to a fluorescent reagent that did not yield a fluorescent hydrolysis product. The fluorescent labeling reagent, Fluoram, proved very valuable in this respect. By combining this reagent with the reduction in ED a single fluorescent peak was obtained. This last improvement was obtained shortly before the end of the tour and technical problems with the detector precluded further fluorescence studies. Therefore, the limits of sensitivity could not be accurately assessed. However, a rough comparison of the chromatographs of the fluorescent-dCMP derivative in Figs. 4 and suggests an improvement of at least several orders of magnitude.

Another section of the project involved the synthesis of standards of modified dNMPs. The modified dNMPs can eventually be employed for several types of experiments. In one, they will be incorporated into reaction mixtures along with the normal nucleotides. These mixtures can then be used to assess the sensitivity of the method. Alternatively, they could serve as standards for dNMPs obtained from digests of DNA from animals or in vitro systems treated with certain potential DNA modifying reagents. Three modified dNMPs were produced. Two from the reactions of chloroacetaldehyde with dCMP and dAMP resp., and one from the reaction of chloral hydrate with dAMP. The products were produced by reactions at 45^0 C. This reduced the reaction time necessary to obtain products. Although it is very likely that the same products will be formed at 37^0 C, it will be necessary to confirm this in the future.

In summary, a high sensitivity fluorescence post-labelling assay, suitable for the detection modified deoxynucleotides obtained from DNA digests has been developed. In addition, several standards of modified deoxynucleotides have been synthesized. In future studies the limits of detection of these standards will be determined, and DNA from animals or in vitro systems (e.g. primary hepatocytes, tissue slices, etc.) treated with chloroacetaldehyde, chloral hydrate or trichlorethylene will be analyzed for the modified nucleotides. In addition, another application of the methods developed here would be to detect new DNA adducts in DNA isolated from animals or in vitro systems treated with genotoxins. Such an application is generally more difficult because standards are not available and therefore the chromatographic region(s) containing the adducts (if any) must be determined. However, such

References

1. Pilot, H.C. (1986) Fundamentals of Oncology. Chemical Carcinogenesis: Mechanisms, **Ed. 3, Marcel Decker (N.Y.)**

2. Boutwell, R.K, Verma, A.K., Ashendel, C.L., and Astrup, E. (1984) Mouse Skin: A Useful Model System for Stydying the Mechanism of Chemical Carcinogenesis. In: Carcinogenesis, **7**, pp. 1-12, Hecker, E., Fusenig, N.E., Kunz, W., Marks, F. and Thielmann, H.W. (eds.) Raven Press, New York.

3. Fearon, E.R. and Vogelstein, B. (1990) A Genetic Model for Colorectal Tumorigenesis. Cell, **61,** 759-767.

4. Hollstein, M.C., Sidransky, D., Vogelstein, B. and Harris, C.C. (1991) p53. Mutations in Human Cancer, Science (Washington, D.C.) **253**, 49-53.

5. Beach, A.C. and Gupta R.C. (1992) Human Biomonitoring and the ^{32}P-postlabeling Assay. Carcinogenesis, **13**, 1053-1074.

6. Azadnia, A., Cambell, R., and Sharma, M. (1994) The Scope of Dansyl vs Fluorescein Label in Fluorescence Postlabeling Assay for DNA Damage, Anal. Biochem., **218,** 444-448.

7. Jain, R. and Sharma, M. (1993) Cancer Res., **53**, 2771-2774.

8. Singer, B. and Grunberger, D. (1983) Molecular Biology of Mutagens and Carcinogens. *Plenum Press, N.Y.*, 15-44.

Figures (On the following pages)

Chromatographs of dNMPs and their reaction products with fluorescent reagents, or with chloroacetaldehyde or chloral hydrate. Injection points are indicated with arrows. New products are marked with a "P". Conditions are indicated in the Methods section.

Fig. 1 dCMP-phosphoramidate mix, unincubated

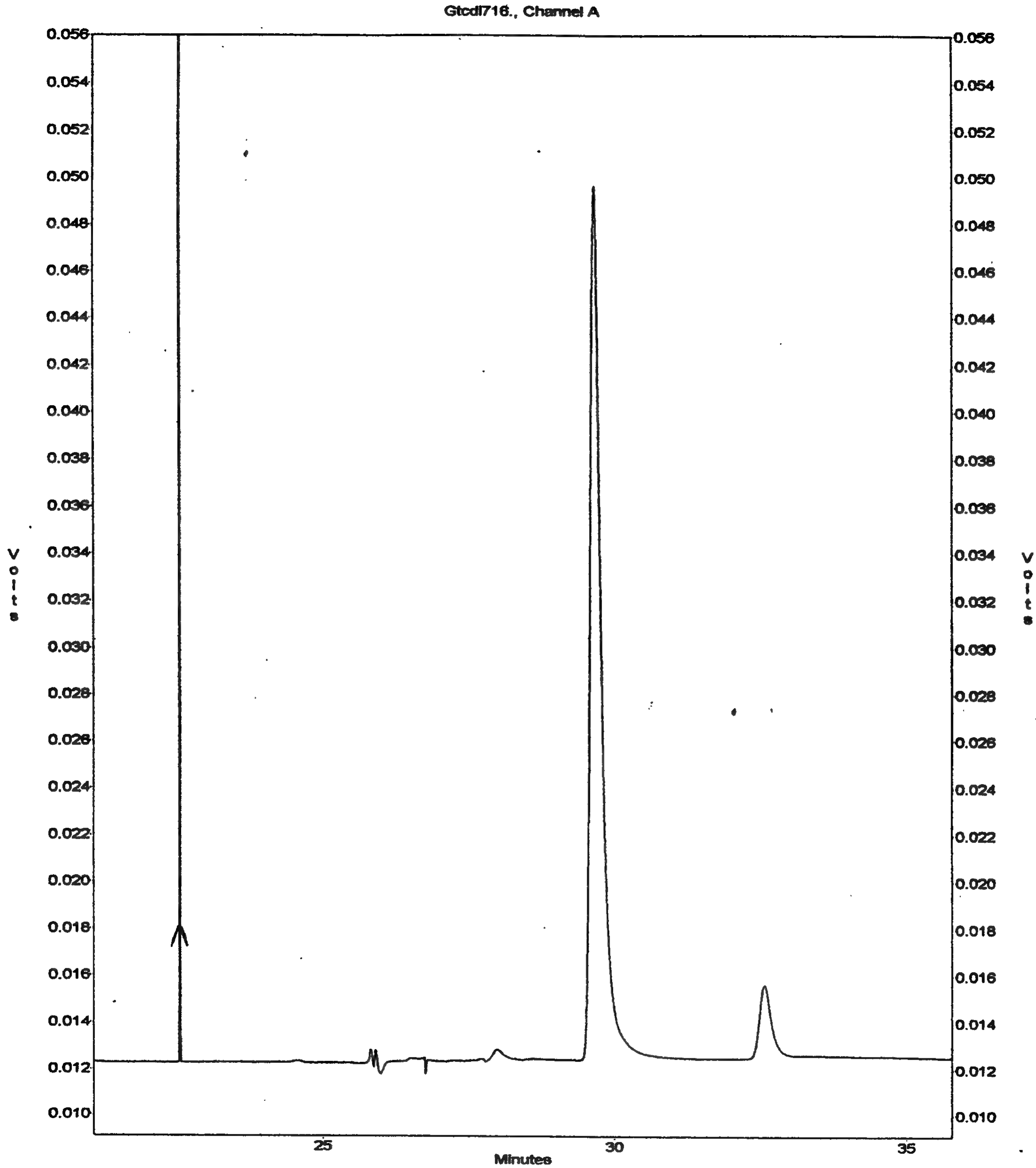


Fig. 2 dCMP - phosphoramidate mix, incubated

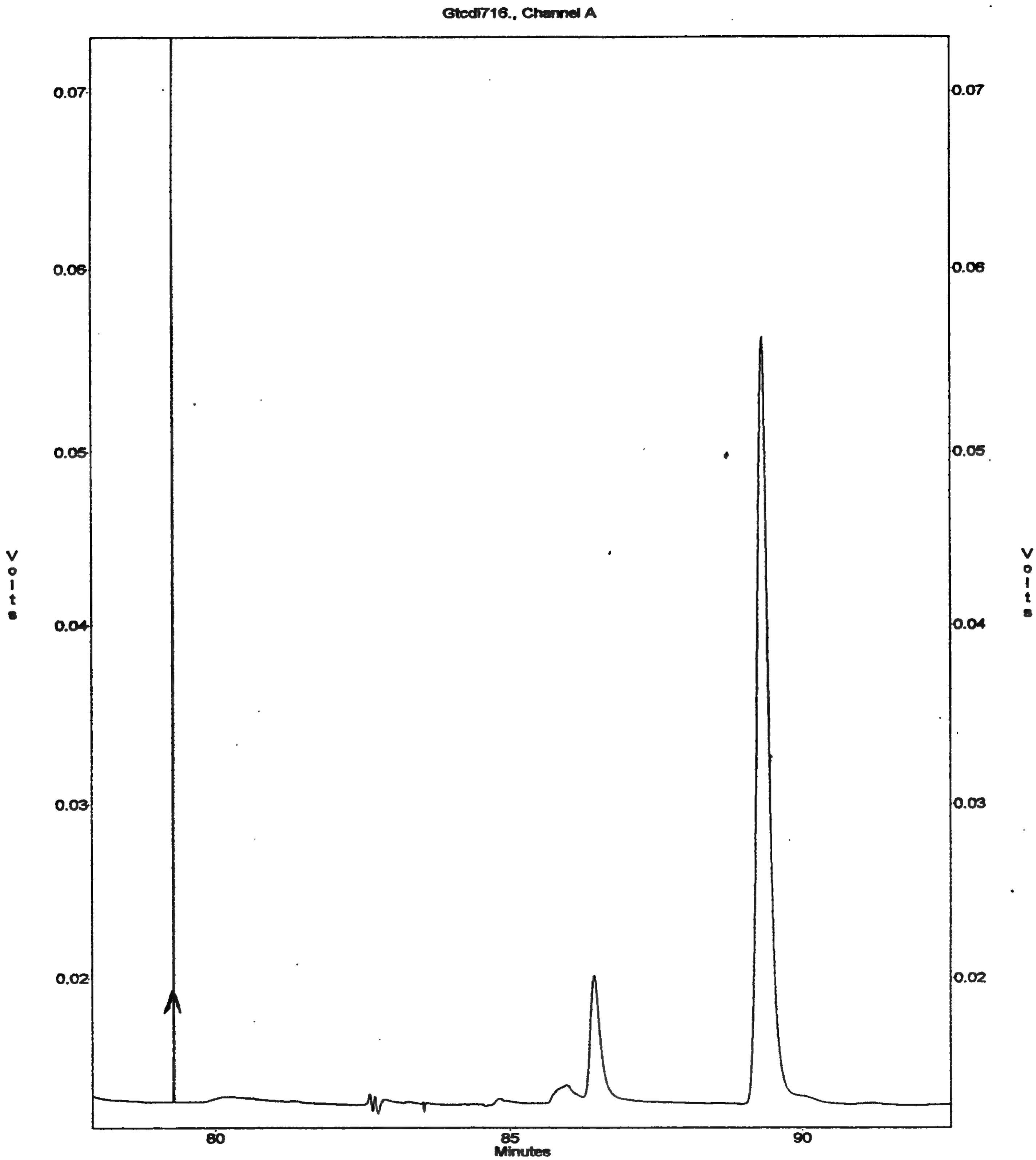


Fig. 3 dCMP - phosphoramidate - Dansyl chloride

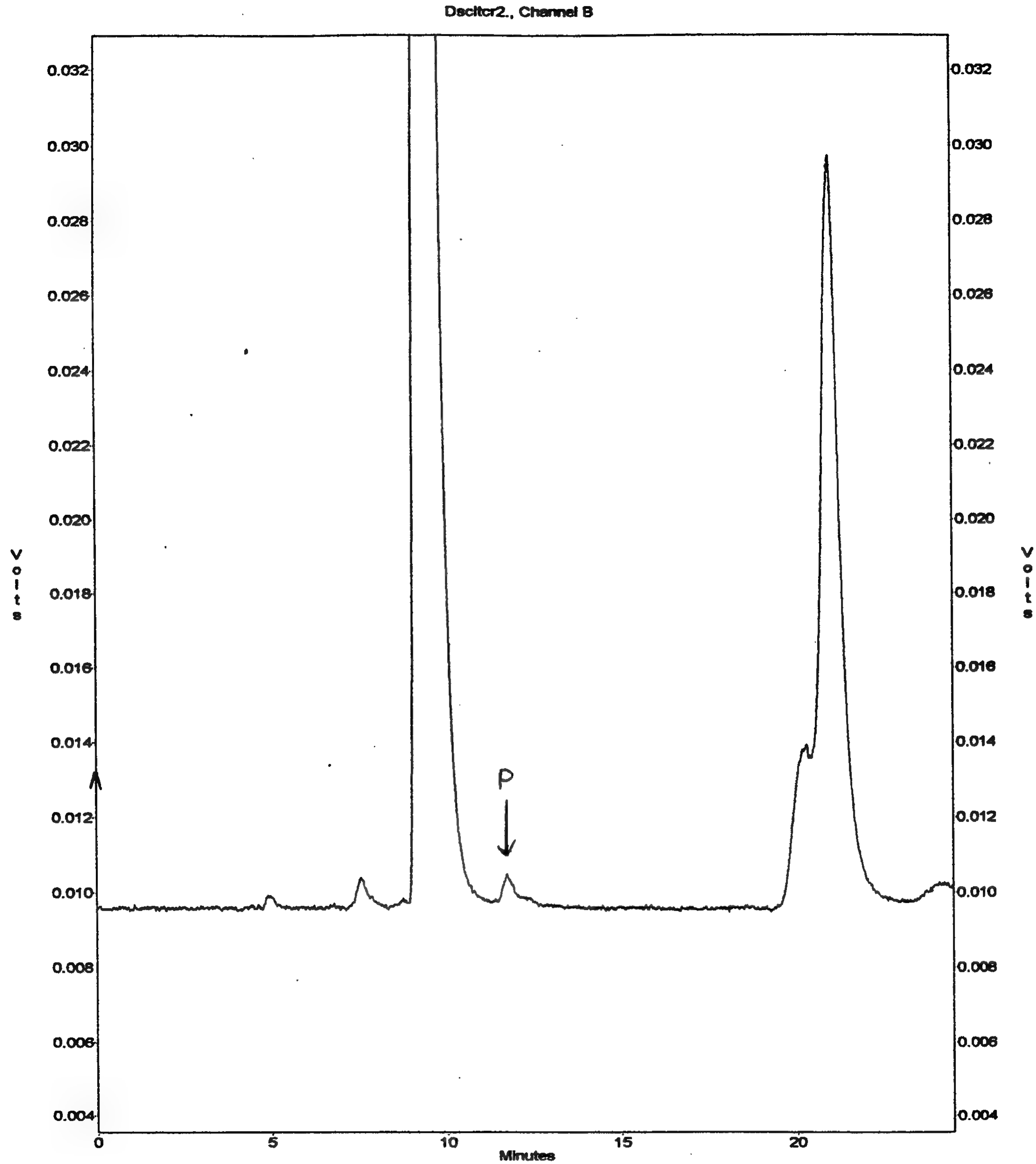


Fig. 4 dCMP - phosphoramidate - Dansyl chloride lyophilized before reaction

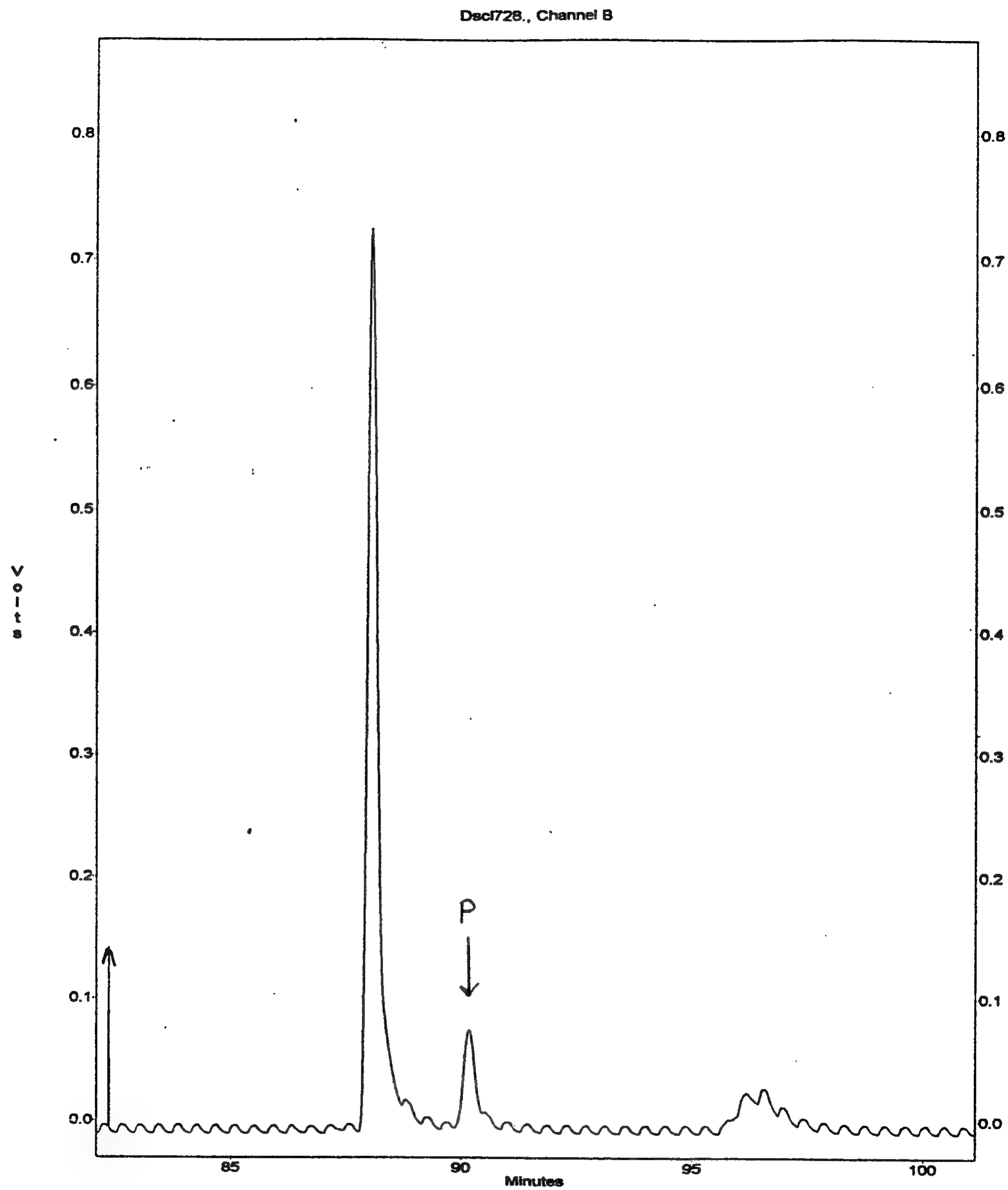


Fig. 5 A. dTMP - phosphoramidate - Fluoram
B. ED - Fluoram

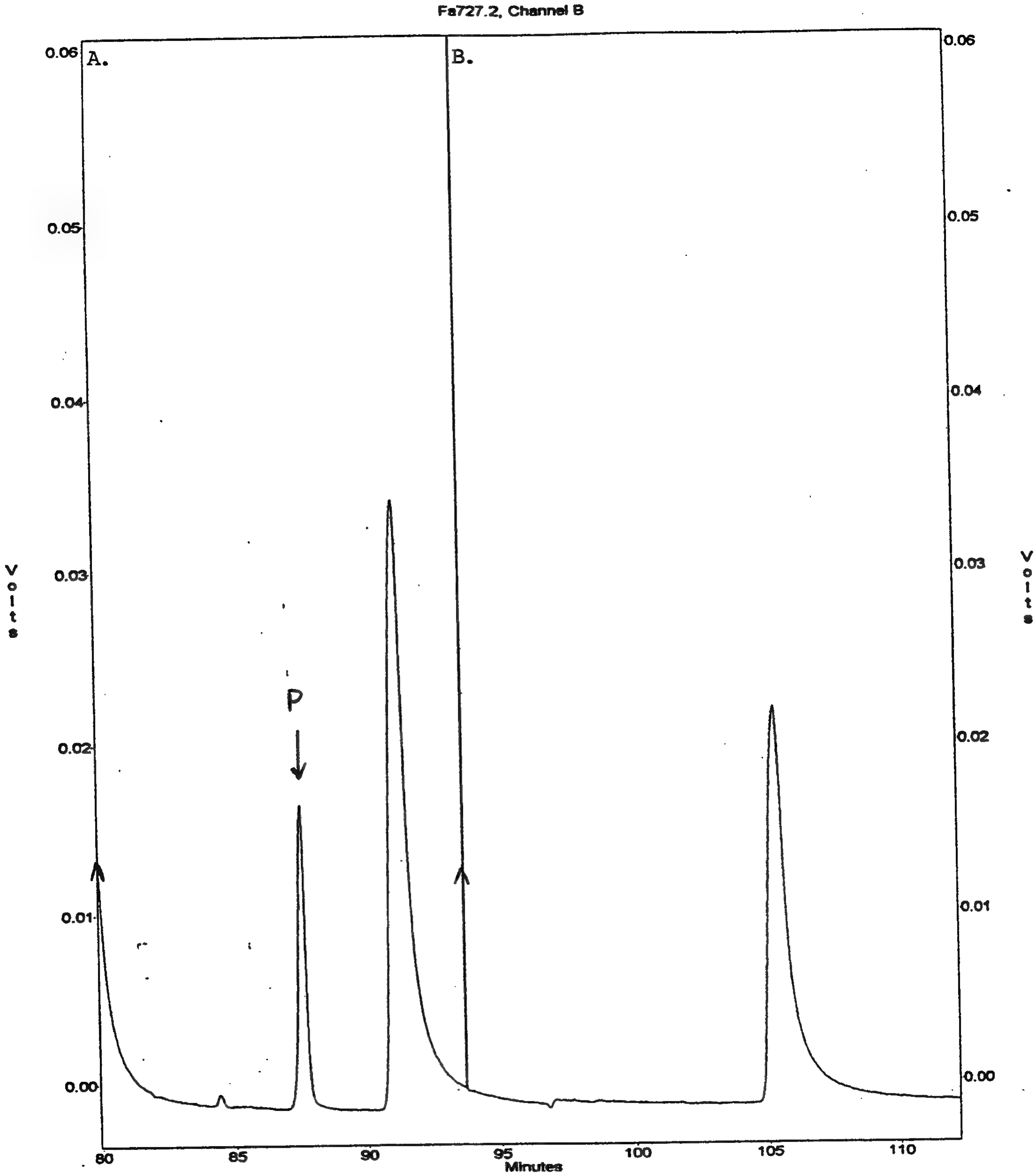


Fig. 6 dCMP - phosphoramidate - Fluoram
lyophilized before reaction

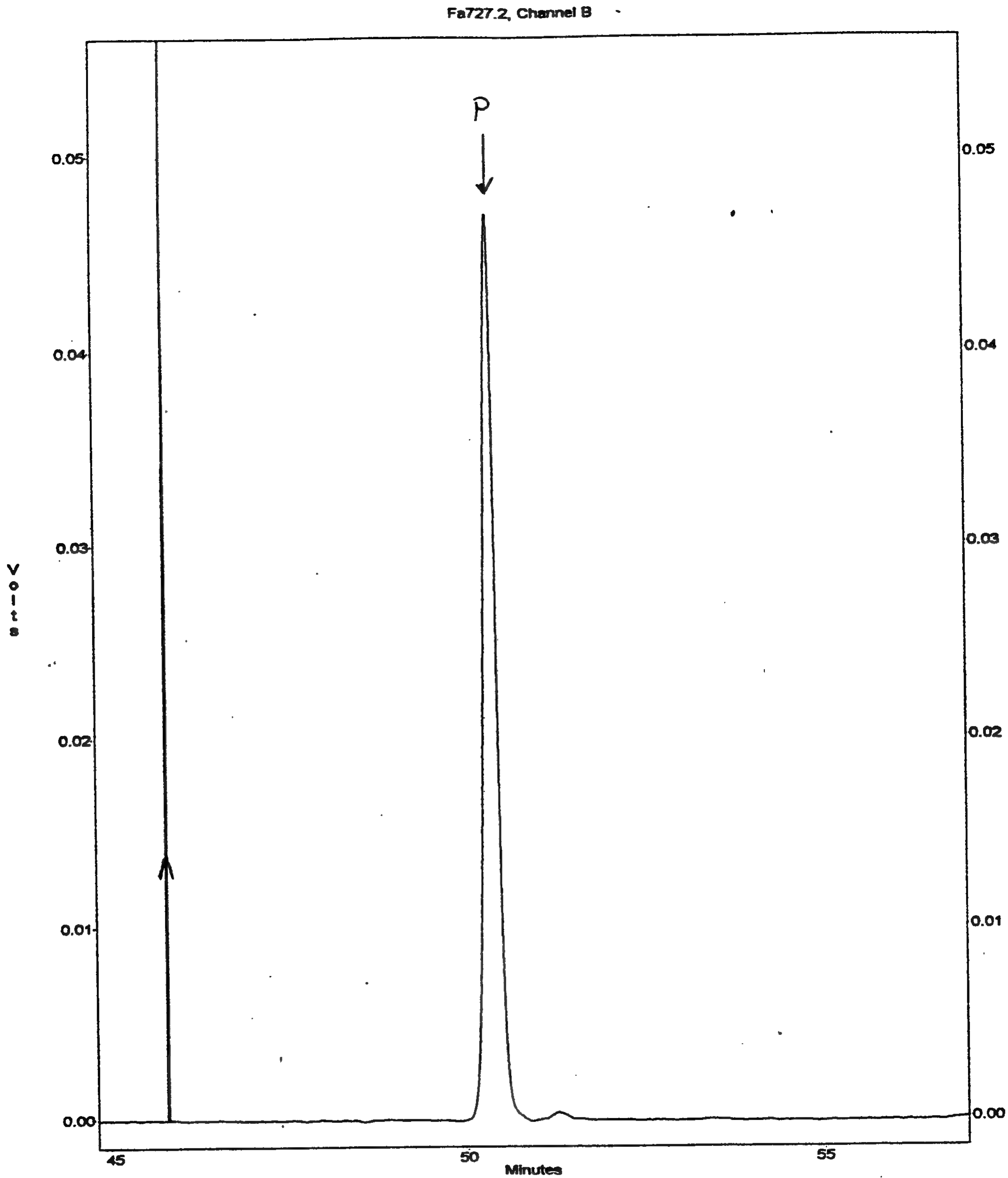


Fig. 7 A. dCMP + chloroacetaldehyde, unincubated
B. chloroacetaldehyde, unincubated

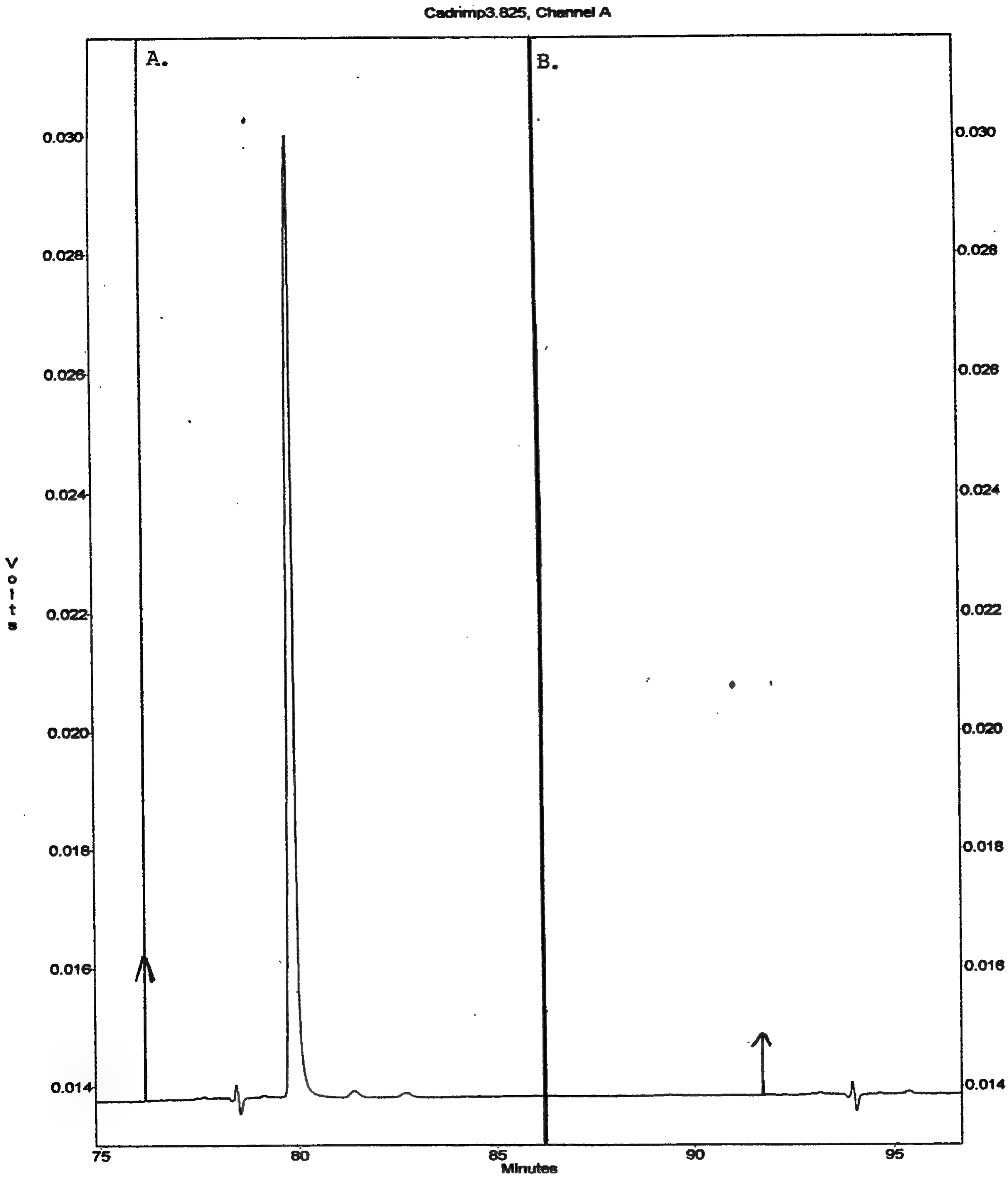


Fig. 8 A. dCMP + chloroacetaldehyde,incubated
B. chloroacetaldehyde, incubated

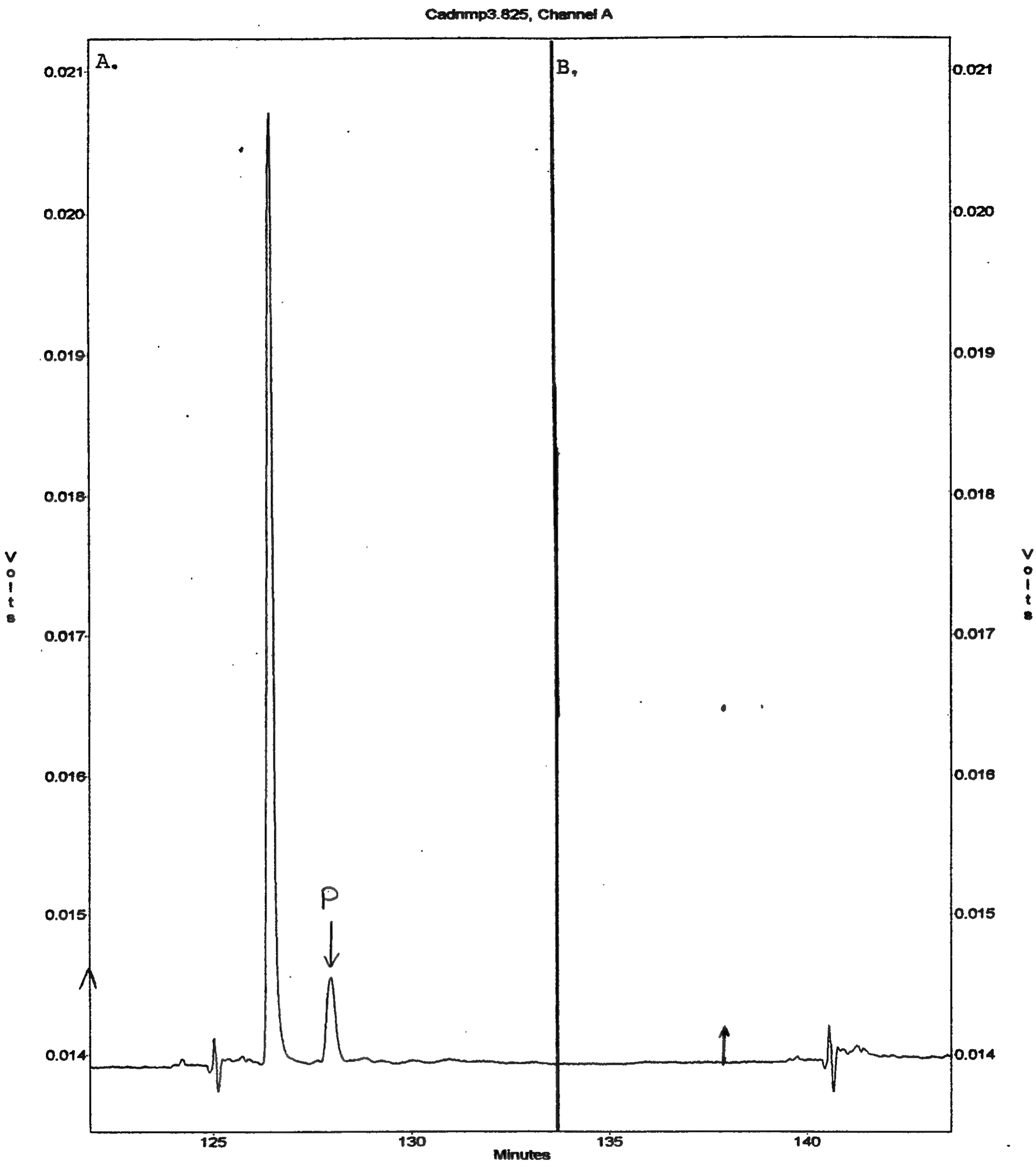


Fig. 9 A. dAMP + chloroacetaldehyde, unincubated
B. dAMP + chloroacetaldehyde, incubated

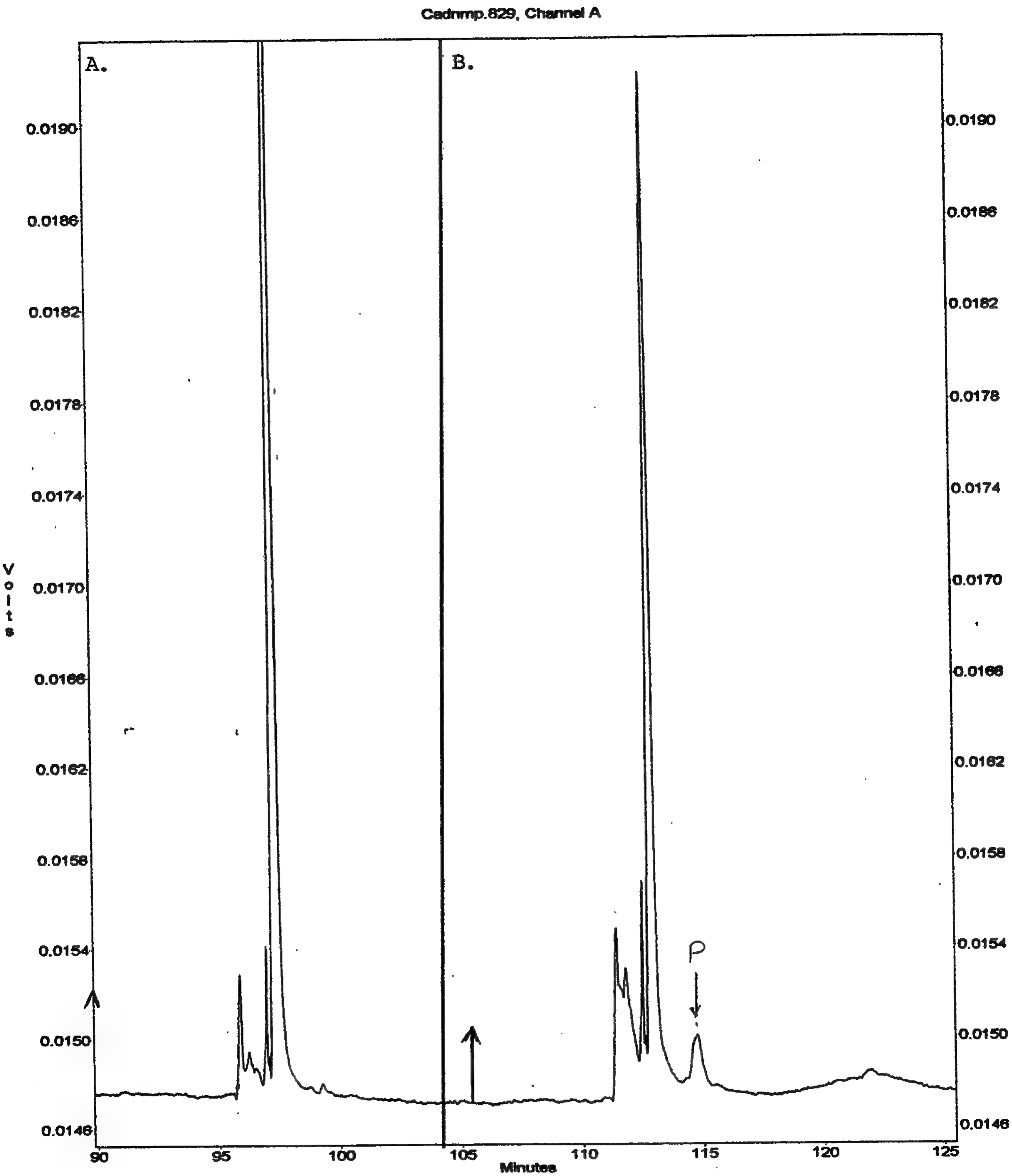
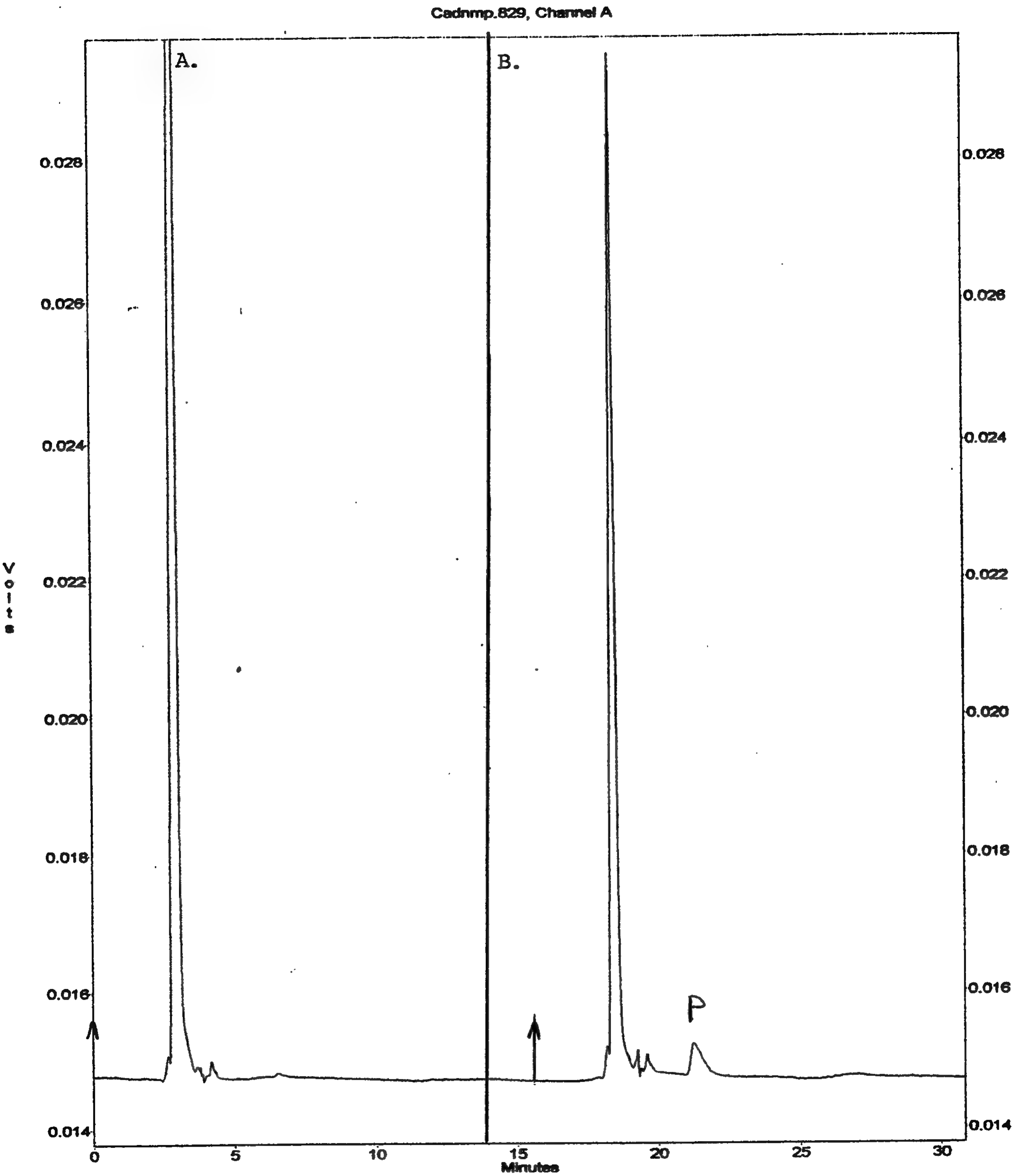


Fig. 10 A. dAMP + chloral hydrate, unincubated
B. dAMP + chloral hydrate, incubated



RELIGIOSITY, SUBJECTIVE HEALTH AND HEALTH BEHAVIOR AMONG A SELECTED RETIRED POPULATION

Glenda Stein Johnson
Assistant Professor
Division of Natural Sciences

Dillard University
2601 Gentilly Boulevard
New Orleans, LA 70122

Final Report for
Summer Faculty Research Program
Armstrong Laboratory

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base
Washington, DC

and

Armstrong Laboratory

August 1995

RELIGIOSITY, SUBJECTIVE HEALTH AND HEALTH BEHAVIOR AMONG A SELECTED RETIRED POPULATION

Glenda Stein Johnson
Assistant Professor
Division of Natural Sciences
Dillard University

Abstract

Health professionals and researchers have begun to consider man's religious practices and beliefs as significant factors in health and as an influence in integrating positive health behavior into individual lifestyles. The relationship of religiosity to subjective health and health protective behaviors were examined in a convenience sample of 58 retired male and female United States Air Force (USAF) members living in two Air Force retirement communities. Only a few significant associations between religiosity and specific health behaviors in this sample were observed. No significant association was found with any of the dimensions of religiosity and subjective health. These findings are discussed in relationship to sample characteristics.

RELIGIOSITY, SUBJECTIVE HEALTH AND HEALTH BEHAVIOR AMONG A SELECTED RETIRED POPULATION

Glenda Stein Johnson

Introduction

Historically, man's religious value system has been a critical aspect of his culture. The scientific community has recently focused considerable attention on the relationship between religiosity and health. Scholars (see review by Levin and Schillier, 1987) examined approximately 250 epidemiologic, sociomedical and biomedical investigations related to religion and health. The reviewers concluded that religion appears to exerts a salutary effect on health. Many of the studies compared the relationship of religious affiliations and health outcomes. Several researchers reported lower mortality rates from smoking-related cancers among Utah Mormons compared to Utah non-Mormons (Lyon, Gardner, & West, 1980) and among California Seventh-Day Adventists compared to California non Seventh-Day Adventists. A lower incidence of breast, cervical, colon and rectal cancers were documented among members of the Mormon and Seventh-Day Adventist Churches (Gardner, 1977, and Phillips, 1979). Mormons and Seventh-Day Adventists had a lower incidence of circulatory diseases (Levin and Vanderpool, 1987, 1989; Armstrong, 1977; and Phillips, 1980); and Mormons had 35% less mortality rate from cardiovascular disease (Lyon, 1978) and greater longevity (Jarvis & Northcott, 1987). It is postulated that the positive health outcomes experienced by these church related groups are due to their strong commitment to health

promoting behavior. Mormon followers adhere to strict health teachings of abstinence from alcohol, tobacco, coffee, tea, habit-forming drugs and conservative sexual and dietary practices. Religious guidelines of Seventh-Day Adventists also proscribe alcohol, tobacco and caffeinated beverages; and discourage the consumption of meats and fat products (Troyer, 1988).

A physical and mental health study of 6,928 adults residing in Alameda County, California was conducted in 1965 (Belloc and et.al, 1971). A 5 1/2-year follow-up investigation (Belloc and Breslow, 1972) revealed that the residents who practiced six or seven basic health-related behaviors were less likely to have died than those who practiced zero to three behaviors. The health behaviors included sleeping 7-8 hours daily; regularity of meals and physical activity; avoiding smoking and drinking and maintaining ideal body weight. A study following 268 persons, aged 60 years and older, for several years, documented an association between exercise; keeping a moderate weight; avoiding smoking and illness and longevity. (Palmore, 1970). With a sample of 401 mothers, (Pratt, 1971) examined a wide range of health practices in relationship between socioeconomic status and health. Mothers practicing good health behavior reported fewer illnesses and a higher level of optimal health.

Religious affiliation and church attendance, used as single measures, have been the most frequently used dimensions of religiosity in health research. Currently, religiosity is viewed

as multidimensional (King, 1967). Other measures frequently used in assessing the multidimensionality of religion are: religious practice, feelings, knowledge and effect (Glock, 1962).

It is conceptualized that religion is linked to health predominantly through prescriptive and proscriptive doctrinal and behavioral expectations which preclude known health risks (Bock, Cochran, and Beeghley, 1987). The theoretical model on which the present research study is based comes in part from the theory of Kenneth Vaux (1976). He postulates that there are certain fundamental religious dispositions that form the basis for those beliefs that influence health. One is the notion of "purity," whereby; one's body is the temple of God and therefore; one is responsible for keeping it healthy.

Second is the concept of "peace in existence." This idea implies that a sense of contentment and purposefulness life comes as a result of an abiding relationship with God as a Friend and a Savior. This sense of well-being may prompt the desire to guard one's health and have regular health evaluations, and is also likely to reduce the amount of stress one experiences when faced with the daily trials of life. A third proposition centers around the idea that immortality (promised to the saved) leads people to either regard and disregard present health. To some, it means immortality of the soul; the shedding of the outer garment or physical body at death. The body can be destroyed, but the soul is immortal. Such a belief can lead individuals to ignore their bodies' health or to practice and extreme ascetic

lifestyle. Some people believe immortality means hope in the resurrection of the body; and they look forward to escape from the terror of death. The Hebraic dread of death leads one to preserve life and cling to vitality.

Research studies with the elderly revealed older adults have established health behavior and strong religious commitment. Religion is especially important to the elderly because it offers consolation and hope during the time when losses, such as youth; health; employment; friends and family are most likely to occur (Moberg, 1990). It has been noted that with advanced aging; total social activity declines, especially church attendance. However, the elderly retain their religious attitudes and beliefs; and perform private religious practices, such as Bible reading and prayer (Mindel and Vaughan, 1978).

There are limited studies investigating the relationship of religiosity to subjective health and behavior among the elderly. The elderly are of particular interest because of their higher morbidity from chronic illnesses and their projected demographic increase in numbers over the next decades.

<u>Methodology</u>

Sample. The association between religiosity and health promoting behavior was examined from data obtained in 1995 from surveys completed by a convenience sample of retired USAF members and dependents, residing in a retirement community. The residents were informed of the study via of the residential newsletter and closed circuit television. Residents who wished

to participate obtained a survey from the office, completed it and returned it to the office personnel. Confidentiality of their responses was assured.

Instrument. A 24-item self-administered questionnaire was used to measure religiosity, subjective health and health promoting behavior. Because of the age and attention span of the subjects, it was necessary for the researcher to develop an instrument which could be completed in approximately 10 minutes. The researcher incorporated ideas from a variety of literature sources in constructing the Religion, Health and Behavior Questionnaire. Four open-ended questions designed to measure the multidemisionality of religiosity were used in this study. They were: (1) "How often did you attend church or participate in religious services?" (2) "How often did you listen to religious service or music on the radio or television?" (3) "How often did you feel close to God?" (4) "How often did you pray?" Questions were asked regarding the activities in the last 12 months and were scaled as frequency (days) per week. Responses to the first question regarding frequency of church attendance were used to measure organizational religiosity, and responses to questions two, three and four measured non-organizational religiosity. Non-organizational religiosity includes personal religious practices, such as prayer and meditation. Subjects were also asked about their church denomination and length of membership.

The questionnaire included a set of questions relative to the individual's lifestyle and perception of health. The dependent

variable, subjective health, was measured according to how the respondents rated their health; ranging from poor to excellent. Five variables were employed for the purpose of assessing health behaviors. They were frequency and amount of alcohol and cigarettes consumption; hours of sleep per day; frequency of exercise per week and the reduction of salt and fat intake in the last 12 months. Social support and the number of traumatic events the respondents experienced in the last 12 months were also included as independent variables. The measure of social support was assessed by asking the respondents if they feel they have enough friends or relatives to talk about private matters or call for help. The subjects also completed questions relative to general demographic information. A focus group, consisting of several volunteer residents, was conducted to evaluate clarity of the pilot form of the instrument. Minor adjustments in the format and directions for completing the questionnaire were made as a result of comments received.

Statistical Methods. The relationships between religious behaviors and health behaviors were explored using Pearson correlation coefficients by gender. Marital status was dummy coded as (1) married, (0) unmarried in the analysis. The relationships between religious behaviors and subject health were first explored using Pearson correlation coefficients, followed by five separate first-order regressions by gender. In each regression, a different one of the control variables is forced-entered prior to religious variables. This analysis will

help rule out competing explanations for a religion/subjective health association. In addition, such an analysis permits the uncovering of any potential "suppressor" effects. It is possible that combination of some or all of the control variables best explains statistically significant associations between religion and subjective health. A regression analysis with more than one control variable in the model was not performed because of the small sample size. Since all the subjects were Caucasians who did not smoke (except for one person), ethnicity and smoking were not considered as controls in the regression analysis.

Results

A total of 27 men with a mean age of 75 years and 31 women with a mean age of 75 years responded to the questionnaire. More than 75% of the sample were over 70 years of age. All the respondents were Caucasian. Seventy-four percent of the sample were married, 7% were single, 3% were divorced and 16% were widowed. Forty-one percent of the sample were dependents of the Air Force retirees. Among Air Force retirees, eighty-five percent had retired ranks as Lieutenant Colonel or Colonel.

Statistics describing the sample are summarized in Table 1. All the participants, except one, rated their health at least in fair condition. Over 85% of the respondents evaluated themselves as in good, very good or excellent health. The distribution of body mass values showed that excess weight was not a serious problem in the sample. Only 7% of the subjects had an unacceptable body mass ratio for their age. Except for one

Table 1: Descriptive Statistics for Data

Variable	Category	Percentage or mean+sd		Median (range)
Age	n=58	74.8+7.2		75 (45,90)
Gender	Male	46.6%	(27/58)	
	Female	53.4%	(31/58)	
Race	Caucasian	100.0%	(58/58)	
Marital Status	Married	74.1%	(43/58)	
	Single/Never Married	6.9%	(4/58)	
	Divorced/Separated	3.4%	(2/58)	
	Widowed	15.5%	(9/58)	
Rank	Col	27.6/%	(16/58)	
	Lt. Col	22.4%	(13/58)	
	Maj	5.2%	(3/58)	
	CWO2	1.7%	(1/58)	
	Dependent	41.4%	(24/58)	
Self-evaluated Health Status				
	excellent	21.4%	(12/56)	
	very good	41.1%	(23/56)	
	good	25.0%	(14/56)	
	fair	10.7%	(6/56)	
	poor	1.8%	(1/56)	
Weight (lb.)	n=57	154.6+28.1		154(105,218)
Height (ft"inches")	n=58	5'6.6"+4.1		5'7"(5'0",6'2")
Body Mass (age adjusted)	low	49.1%	(28/57	
	normal	43.9%	(25/57)	
	high	7.0%	(4/57)	

Table 1: Descriptive Statistics for Data (continued)

Variable	Category	Percentage or mean+sd		Median (range)
Length of membership (years)		54.7+24.1		62(1,90)
Smoke	yes	1.7%	(1/58)	
	no	98.3%	(57/58)	
Cigarettes (per/day)	n=1	10		10(10,10)
Consume alcohol	yes	53.4%	(31/58)	
	no	46.6%	(27/58)	
Drinks (per/day)	n=30	3.6+6.7		0.5(0.30)
Sleep (hours/day)		7.6+1.1		8(5,10)
	5	1.8%	(1/57)	
	6	17.5%	(10/57)	
	7	26.3%	(15/57)	
	8	36.8%	(21/57)	
	9	12.3%	(7/57)	
	10	5.3%	(3/57)	
Exercise (days/week)		4.1+2.3		4(0,7)
	0	15.5%	(9/58)	
	1	1.7%	(1/58)	
	2	3.4%	(2/58)	
	3	15.5%	(9/58)	
	4	19.0%	(11/58)	
	5	12.1%	(7/58)	
	6	15.5%	(9/58)	
Eliminate salt	yes	72.4%	(42/58)	
	no	27.6%	(16/58)	
Eliminate fat	yes	82.8%	(48/58)	
	no	17.2%	(10/58)	
Close friend	yes	86.2%	(50/58)	
	no	13.8%	(8/58)	
# of Traumatic events (in the past two years)	0	27.6%	(16/58)	
	1	20.7%	(12/58)	
	2	25.9%	(15/58)	
	3	15.5%	(9/58)	
	4	1.7%	(1/58)	
	5	3.4%	(2/58)	
	6	1.7%	(1/58)	
	8	1.7%	(1/58)	
	10	1.7%	(1/58)	

Table 1: Descriptive Statistics for Data (continued)

Variable	Category	Percentage or mean+sd		Median
Church attendance (days/week)		1.5+1.4		1(0,7)
	0	12.1%	(7/58)	
	1	51.7%	(30/58)	
	2	20.7%	(12/58)	
	3	8.6%	(5/58)	
	4	1.7%	(1/58)	
	7	3.4%	(2/58)	
	occasionally	1.7%	(1/58)	
Listen (days/week)		1.8+2.4		1(0,7)
	0	33.3%	(19/57)	
	1	24.6%	(14/57)	
	2	3.5%	(2/57)	
	3	5.3%	(3/57)	
	4	1.8%	(1/57)	
	5	5.3%	(3/57)	
	7	10.5%	(7/57)	
	occasionally	15.8%	(9/57)	
Close to God (days/week)		5.6+2.4		7(0,7)
	0	5.3%	(3/57)	
	1	5.3%	(3/57)	
	2	8.8%	(5/57)	
	3	1.8%	(1/57)	
	4	1.8%	(1/57)	
	5	5.3%	(3/57)	
	6	1.8%	(1/57)	
	7	70.2%	(40/57)	
Pray to God (days/week)		5.8+2.3		7(0,7)
	0	5.3%	(3/57)	
	1	3.5%	(2/57)	
	2	8.8%	(5/57)	
	4	3.5%	(2/57)	
	5	1.8%	(1/57)	
	7	77.2%	(44/57)	
Church Denomination				
	Baptist	5.5%	(9/58)	
	Methodist	0.3%	(6/58)	
	Mormon/Jehovah W	1.7%	(1/58)	
	Catholic	19.0%	(11/58)	
	Nondenom Prot	13.8%	(8/58)	
	Other	32.8%	(10/58)	
	Not attend church	6.9%	(4/58)	

subject, none of the participants had smoked in the past 12 months. The average frequency of exercise was four days per week and 80% of the sample exercised at least twice a week. About half of the sample had consumed alcohol in the past 12 months. The amount ranged from 5 to 30 drinks a week. Most of the subjects (54%) had slept least 8 hours everyday, with an average of 7.6 hours. The majority of the sample had decreased their salt or fat intake, 72% and 83% respectively. Eighty-six percent of the sample felt they had enough friends or relatives for private matters or help. Most of the participants (72%) experienced traumatic life events in the last two years.

High levels of religiosity were found in the sample. Eighty-six percent of the sample attended church or participated in religious related services at least once a week. Except for four subjects (7%), everyone in the sample was a church member, with a median membership of 62 years. About half (51%) of the sample listened to religious services or music on the radio or television at least once a week. Seventy percent of the sample felt close to God everyday, while 77% of the sample prayed to God everyday. All the sample, except two subjects (3%) participated in at least one form of non-organizational religious activities at least once a week.

Table 2 is the correlation matrix of the predictors (religion and control variables) and the outcomes (subjective health and health behaviors) for all the subjects and for men and women separately. As shown in Table 2, neither church attendance nor

Table 2: Correlation Matrix (Pearson Correlation) for all subjects and by gender. (first row: men; second row: women; third row: all) (continued)

		Subjective Health	Drink	Sleep	Exercise	Eliminate Salt	Eliminate Fat
Organizational Religiosity (church attendance)	men	-0.1	0.04	-0.02	-0.47	0.17	0.23
	women	0.25	0.63*	0.23	0.13	0.01	0.03
	all	0.08	0.31*	0.09	-0.15	0.10	0.15
Non-organizational Religiosity (average of the following three items)	men	-0.30	-0.18	-0.29	-0.14	0.03	0.26
	women	0.04	-0.13	-0.31	0.19	-0.30	-0.10
	all	-0.18	-0.17	-0.34*	0.00	-0.07	0.15
Listen/Religious Service	men	0.03	0.10	-0.10	0.11	-0.23	-0.08
	women	0.17	0.05	-0.05	0.25	-0.51*	-0.20
	all	0.09	0.06	-0.10	0.17	-0.32*	-0.12
Close to God	men	-0.35	-0.03	-0.08	-0.07	0.17	0.24
	women	0.14	0.01	-0.28	0.05	0.05	-0.05
	all	-0.16	-0.02	-0.22	-0.04	0.15	0.16
Pray	men	-0.17	-0.37	-0.37	-0.23	0.12	0.31
	women	-0.27	-0.50*	-0.30	0.07	-0.06	-0.14
	all	-0.22	-0.42*	-0.38*	-0.10	0.08	0.17
Length of Membership	men	-0.37	-0.14	0.05	-0.19	-0.09	0.02
	women	-0.17	-0.33	-0.06	-0.21	0.02	-0.26
	all	-0.28	-0.23	-0.03	-0.21	-0.02	-0.09

Continued

Table 2: Correlation Matrix (Pearson Correlation) for all subjects and by gender. (first row: men; second row: women; third row: all)

		Subjective Health	Drink	Sleep	Exercise	Eliminate Salt	Eliminate Fat
Gender	men	-	-	-	-	-	-
	women	-	-	-	-	-	-
	all	0.15	0.05	0.21	0.08	-0.12	-0.12
Age	men	-0.10	-0.02	0.39*	-0.22	-0.15	-0.05
	women	0.02	0.03	-0.06	-0.02	-0.01	-0.05
	all	-0.04	0.00	0.19	-0.11	-0.09	-0.05
Marital Status	men	-0.20	0.16	-0.17	-0.37	0.10	0.19
	women	-0.14	0.07	0.41*	0.09	0.01	-0.33
	all	-0.08	0.11	0.26*	0.00	-0.01	-0.17
Close Friend (social support)	men	-0.07	0.23	0.29	0.10	0.15	0.03
	women	0.32	0.15	0.03	0.22	0.25	-0.15
	all	0.13	0.19	0.16	0.16	0.20	0.05
# Traumatic Events	men	-0.51*	-0.18	-0.03	-0.14	-0.35	-0.23
	women	-0.19	0.16	-0.17	0.12	0.38*	0.04
	all	-0.29*	0.03	-0.17	0.03	0.17	-0.01

*p 0.05

any of the variables of non-organizational religiosity were correlated with subjective health. However, some significant associations between religiosity and health behaviors were observed. Church attendance was positively correlated with consumption of drinks for females and for all subjects ($r=0.63$ and $r=0.31$, respectively). Negative association was observed between exercising and church attendance in men ($r=0.47$). The duration of sleep everyday clearly had the largest negative correlation to non-organizational religiosity ($r=-0.34$). Among the variables of non-organizational religiosity, listening to religious service and music was negatively correlated to decreasing salt intake and praying was negatively correlated to both consumption of alcoholic drinks and sleep duration.
An examination of the correlation matrix through scatter plots of the data showed that most of the significant correlation coefficients were due to very few influential observations.

Education and income are frequently collected as controls in this type of study. Retired rank instead of education and income were collected in this study however, they were found not useful in the analysis due to the fact that many participants were dependents of the retirees; and information regarding their occupation, education or income was not collected.

Discussion

Although the null hypotheses, "religiosity is not significantly associated with positive health behavior" and "religiosity is not significantly associated with subjective

health," guided this investigation; the findings failed to reject these suppositions. As shown in Table 1, a majority of the elderly in this sample frequently engaged in both organizational and non-organizational religious activities. Most of them were healthy subjectively and indicated they practiced at least six health protecting behaviors. These characteristics were similar to the 268 elderly adults in the follow-up Alameda Health Study (Palmore, 1970).

Several factors may have caused the failure of this study to document significant associations among the variables. One was homogeneity of the sample relative to their health status, health and religious behaviors; making it difficult to reveal the relationships among them. Second was sample size. Since the number of non-religious and less healthy persons were disproportionately low, a larger sample may have been needed to detect relationships. The third factor is self-selection bias. Participation in the study, required the subjects to obtain and return a completed questionnaire to the office personnel. People who were not religiously active might not be interested in completing the questionnaires and residents who were ill or had limited mobility would possibly be unable to participate. Therefore, the target population might not be well represented by the sample. Lastly, a more sensitive instrument but easily self-administered by the elderly appear warranted. Longitudinal studies may be needed to examine more accurately the relationship of religiosity, health and health behavior among any age group.

Conclusion

It is well documented that medical costs are accelerating and many persons lack health insurance. It is also generally recognized that traditional health education used to alter personal lifestyle has experienced limited success. More diverse approaches and mechanisms must be developed to assist with current health education efforts. The literature reports that 94% of the general public believe in God or a Universal spirit who functions as God (Glock, 1962). Integrating religion and its value system into the life of some individuals, may be the key in motivating important lifestyle changes. Basic spiritual principles and Biblical scriptures can be incorporated into coherent health programs offered in schools, workplaces, clinical settings and the military. An example is: "the body is intended to be treated as a temple and caring for the body with a good diet, ample exercise and etc. is a way to worship one's creator." This approach can be used with persons who are deeply spirited, such as the elderly; and those whose spirituality is less developed. The discussions would avoid specific religious denominational doctrine.

Health professionals can be trained to include measures of religious behavior in performing health assessments. This information will then be used to maximize aspects of the client's belief system that promote positive health behavior. Including religion in the health care process, also has the potential for generating greater compliance with other medical plans.

BIBLIOGRAPHY

Armstrong, B., Van Merwyk, A.J., & Coates, H., Blood pressure in Seventh-day Adventist Vegetarians. (1977) American Journal of Epidemiology, 105, 444-449.

Belloc, N., Breslow, L. and Hochstim, J. Measurement of physical health in a general population survey. (1971) American Journal of Epidemiology, 93, 328-336.

Belloc, N.B. and Breslow, L. Relationship of physical health status and health practice. (1972) Preventive Medicine 1, 409-421.

Boch, E.W., Cochran, J.K., & Beeghley, L., Moral messages: The relative influence of denomination on the religiosity-alcohol relationship, (1987) The Sociological Quarterly, 28, 1, 89-103.

Cullman, O., Immortality of the soul or resurrection of the dead, (1965) In Immortality and Resurrection, ed. K. Stendahl, New York: Macmillan.

Gardner, J.W., & Lyon, J.L., Low evidence of cervical cancer in Utah. (1977) Gynecological Oncology, 5, 68-80.

Glock, C.Y., On the study of religious commitment. (1962) July-Aug, Religious Education Research Supplement.

Jarvis, G.K., & Northcott, H.C., Religion and differences in morbidity and mortality. (1987) Social Science and Medicine, 25, 813-824.

King, M, Measuring the religious variable. Nine proposed dimensions. (1967) Journal for the Scientific Study of Religion, 6:173-85.

Levin, J.S. and Schiller, P.L. Is there a religious factor in health? (1987) Journal Religion and Health, 26, 9-36.

Levin, J.S., & Vanderpool, H.Y. Is religion therapeutically significant for hypertension? (1989) Social Science and Medicine, 29, 69-78.

Levin, J.S. & Vanderpool, H.Y. Is frequent religious attendance really conducive to better health?: Toward an epidemiology of religion. (1987) Social Science and Medicine, 24, 589-600.

Lyon, J.L., Gardner, J.W., & West, D.W., Cancer incidence in Mormons and non-Mormons in Utah, 1967-1975, (1980) Journal of the National Cancer Institute, 65, 1055-1061.

Lyon, J.L., Wetzler, H.P., Gardner, J.W., Klauber, M.R., & Williams, R.R., Cardiovascular mortality in Mormons and non-Mormons in Utah, 1968-1971, (1978) American Journal of Epidemiology, 108, 357-366.

Mindel, C.H. and C.E. Vaughan, A Multidimensional approach to religiosity and Disengagement. (1989) Journal of Gerontology, 33:103-106.

Moberg, D.O. Religion and Aging. In Kenneth Farraro, ed. Gerontology: Perspectives and Issues. N.Y.: Springer.

Palmore, E. Health practices and illness among the aged. The Gerontologist. (1970) 10:4, 313-316.

Phillips, R.L, Role of lifestyle and dietary habits in risk of cancer among Seventh-day Adventists, (1979) Cancer Research, 35 (11), 3513-3522.

Phillips, R.L., Kuzma, J.W., Beeson, W.L., & Lotz, T., Influence of selection versus lifestyle on risk of fatal cancer and cardiovascular disease among Seventh-day Adventists, (1990) American Journal of Epidemiology, 112 (2), 296-314.

Pratt, L. The relationship of socioeconomic status to health. (1971) American Journal of Public Health 61, 281-291.

Troyer, H. Review of cancer among four religious sects: Evidence that life-styles are distinctive sets of risk factors. (1988) Social Science and Medicine, 26, 1007-1017.

Vaux, Kenneth, Religion and health, (1976) Preventive Medicine, 5, 522-536.

ACKNOWLEDGEMENTS

The author extends special thanks to Ms. Susan Chao for performing the statistical analysis and offering other helpful suggesions regarding this research.

Manfred Koch report not available at time of publication.

L. Krishsnamurthy Report not available at time of publication.

The Check Mark Pattern

Pandu Kulkarni
Associate Professor
Department of Mathematics and Statistics

University of South Alabama
Mobile, AL 36688

Final Report for:
Summer Faculty Research Program
Armstrong Laboratory

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base
Washington, D.C.

and

Armstrong Laboratory

August 1995

The Check Mark Pattern

Pandu Kulkarni, PhD
Associate Professor of Statistics
Department of Mathematics and Statistics
University of South Alabama

ABSTRACT

We exhibit two studies (one epidemiological and one clinical), both with apparently paradoxical findings characterized by group (index versus control) similarity on the dependent (health) variable (Y) means, a significant group difference on the independent variable (X) means (index mean greater than the control mean) and a positive correlation between Y and X in the index group, causing index subjects with low values of X to have a lower Y mean than the controls and index subjects having high values of X to have a higher Y mean than the controls. This pattern has been called the "check mark" pattern. We predict this pattern using a linear model and use the model to estimate exposure effects in the epidemiologic study. Additionally, we show that a previously published study of the check mark pattern suggesting reverse causation is incorrect.

The Check Mark Pattern

Pandu M. Kulkarni

1. INTRODUCTION

We consider medical studies of a health measure (Y) and an independent variable (X) in patients classified to index and control groups. In clinical studies the index subjects are generally "cases", who have been diagnosed as having a disease or syndrome and in epidemiologic studies of health and exposure to some potentially harmful substance, the index subjects are identified as having been "exposed" to the substance. In the epidemiolgical study, X is a biomarker for exposure and in the clinical study X is a measure of disease or syndrome severity. In both types of studies we assume that the mean value of X in controls is less than the mean value of X in index subjects. Thus, if index and control subjects are otherwise similar one would expect that index subjects with low values of X should have a similar Y mean as control subjects and, if X is positively correlated with Y in the index group, one would expect that index subjects with large values of X would have a larger Y mean than controls. In fact, because index subjects have a larger mean value of X, one would further expect that the overall index Y mean would be greater than the overall control Y mean. In this paper, we consider studies where this expected pattern is not observed. The observed pattern is one in which the index and control means on Y are nearly equal and X and Y are positively correlated in the index group. In this situation, the Y mean among index subjects with low values of X will be less than the control Y mean and the Y mean among index subjects with high values of X will be greater than the control Y mean. Two examples of this "check mark" pattern

are presented, modeled and interpreted. The first example arises in an epidemiological study and the second in a clinical study.

There is disagreement regarding the interpretation of the check mark pattern in epidemiological studies. A recent study of the pattern concluded that the pattern is suggestive of reverse causation, in which Y causes changes in X rather than X causing changes in Y (Flanders et al 1992). The Flanders paper has been cited by the Institute of Medicine (1994) in their recent review of Agent Orange research. We show that Flanders' arguments are incorrect and give alternative interpretations for the pattern.

The purpose of this paper is to describe the check mark pattern, both empirically (in Section 2) and with statistical models (in Section 3), to use the models to estimate exposure effects in the epidemiological example, to explore reverse order causation in the context of our models and to indicate an error in the reverse order causation argument of Flanders et al (1992). We do not intend to provide a statistical analysis of these data, as they have already been analyzed elsewhere in rigorous detail.

2. EXAMPLES

Example 2.1

The Air Force Health Study (AFHS) is a 20-year prospective epidemiological study of possible health effects from exposure to herbicides and their contaminant, 2,3,7,8-tetrachlorodibenzo-p-dioxin (dioxin) in veterans of Operation Ranch Hand, the unit responsible for spraying Agent Orange and other herbicides in Vietnam from 1962 to 1971

(Wolfe et al 1990). A group of comparison veterans who served in Southeast Asia during the same period but who were not involved with spraying herbicides serve as controls. Comparison veterans were matched to Ranch Hand veterans on age, military occupation and race. The study compares the health, reproductive outcomes and mortality of Ranch Hands veterans with comparison veterans. Physical examinations were administered in 1982, 1985, 1987 and 1992. Examinations are planned for 1997 and 2002. Since 1987, a measurement of dioxin in serum has been used as the index of exposure (Wolfe et al 1992).

Nine hundred fifty two Ranch Hands and 1,281 Comparisons attended the 1992 AFHS physical examination of whom 984 Ranch Hands and 1,084 Comparisons had quantitatable dioxin results (Grubbs et al 1995). An investigation of endocrine function included serum insulin, measured in mIU/ml. Because we were interested in the association between dioxin (X) and insulin (Y) in nondiabetic subjects, we excluded known diabetics and subjects with 2 hour post-prandial glucose levels greater than 200 mg/ml (125 Controls and 115 Ranch Hands), leaving 959 controls and 779 Ranch Hands with insulin and dioxin levels for consideration.

We studied the relation between insulin and dioxin after both had been logarithmically transformed (the number 1 was added to the dioxin result before taking the logarithm) using linear regression and t-tests. The data is shown in Figure 1, with a least-squares line overlaid on the Ranch Hand data. The coefficient of log(dioxin+1) in the

regression of log(insulin) in Ranch Hands is statistically significant (slope=0.18, standard error=0.031).

INSERT FIGURE 1 HERE

Table 1 gives summary statistics by group.

Table 1

Log(Dioxin+1) and Log(Insulin) by Group

Group	N	Log(dioxin+1) Mean (Std Dev)	Log(insulin) Mean (Std Dev)
Ranch Hand	779	2.7 (0.99)	4.24 (0.88)
Control	959	1.8 (0.44)	4.25 (0.83)

The group log(insulin) means were not significantly different (mean difference=-0.01, 95% CI: -0.07, 0.09).

We stratified the data to four levels according to group and dioxin level. The first stratum was comprised of control subjects with dioxin less than 10 ppt, the value we regard as the threshold for background exposure. The remaining three strata, named "Low", "Medium" and "High" were determined by the cut points 10 ppt and 24 ppt in the dioxin distribution in Ranch Hands. Table 2 gives summary statistics for log(insulin) and log(dioxin+1) in each of these four strata.

Table 2

Log(Dioxin+1) and Log(Insulin) by Dioxin Category

Stratum	N	Log(dioxin+1) Mean (Std Dev)	Log(insulin) Mean (Std Dev)
Control	942	1.5 (0.55)	4.2 (0.83)
Low	341	1.8 (0.44)	4.1 (0.84)
Medium	221	2.8 (0.24)	4.3 (0.87)
High	217	4.0 (0.59)	4.4 (0.90)

The data is summarized in Figure 2 by dioxin category, showing the check mark pattern.

Example 2.2

A recent study of excitatory amino acids in cerebral spinal fluid (CSF) and pain, measured by the tender point index (TPI) (Russell et al 1994) in patients with fibromyalgia, included 41 patients diagnosed as having fibromyalgia and 37 matched controls (matched on age and ethnicity); we call this study the CSF excitatory amino acid (CSF-EAA) study.

In the CSF-EAA study cerebral spinal fluid (CSF) samples were obtained in 41 patients with fibromyalgia syndrome (FS) and 37 normal controls (C) by standard lumbar puncture. Taurine (Y) was measured in CSF as previously described using specific radioimmunoassays (Giovengo et al 1995) and was reported in ng/µl. The severity of

tenderness (X) was measured by the tender point index (TPI), which involved digital palpation at 18 specific soft tissue sites (Russell et al 1986). In addition to inquiring about pain induced by palpation, the physician closely observed the patient for any physical response and scored each site as follows: 0 (no pain), 1 (tenderness reported but without physical response), 2 ("semi-objective" tenderness demonstrated by a physical response such as a wince or withdrawal), 3 (very exaggerated physical response) and 4 (untouchable; patient will not allow palpation at a given site fearing unbearable pain). The diagnosis of FS was based on the continuous presence of musculoskeletal pain for 3 months and "semi-objective" tenderness at 11 or more of the 18 tender points (Wolfe et al 1990). The data is shown in Figure 3, with a least-squares line overlaid on the FS data. The coefficient of TPI in the regression of taurine in FS patients is statistically significant (slope=0.12, standard error=0.031).

INSERT FIGURE 3 HERE

Table 3 gives summary statistics by group.

Table 3

Tender Point Index and Taurine by Group

Group	**N**	**Tender Point Index Mean (Std Dev)**	**Taurine Mean (Std Dev)**
Fibromyalgia	41	34.9 (11.89)	6.6 (2.75)
Control	37	0.4 (0.98)	5.9 (2.61)

The mean value of taurine did not vary significantly with group (mean difference=1.13 ng/μl, 95% CI: -0.06 ng/μl , 2.32 ng/μl). We stratified the data to four levels according to group and TPI level. The first stratum was comprised of control subjects. The remaining three strata, named "Low", "Medium" and "High" were determined by the tertiles (31 and 40) of the TPI distribution in FS patients. Table 4 gives summary statistics for taurine and the TPI in each of these four strata.

Table 4

The Tender Point Index and Taurine by TPI Category

Stratum	N	Tender Point Index Mean (Std Dev)	Taurine (ng/μl) Mean (Std Dev)
Control	37	0.35 (0.98)	5.88 (2.62)
Low	15	22.3 (5.62)	5.29 (2.26)
Medium	14	36.1 (2.43)	6.58 (1.80)
High	12	49.4 (4.21)	8.16 (3.49)

The data is summarized in Figure 4 by TPI category, showing the check mark pattern.

3. MODELING THE CHECK MARK PATTERN

3.1 Definition of the Check mark Pattern

The patterns in each of these two examples are caused by nearly equal group means on Y and a positive correlation between X and Y in the index group. The conditional means shown in Tables 2 and 4 then exhibit the pattern, with the control mean on the health variable (Y) being greater than the index mean in the Low category and an increasing sequence of conditional means on Y in the index categories (Low, Medium and High) determined by two cut points (c_{11} and c_{21}) in the distribution of the independent variable (X) in the index group, where c_{11} and c_{12} are approximate tertiles of X in the index group. The control stratum is comprised of control subjects, possibly truncated with X at or below a cut point c_{10}, which we assume as a high quantile of the distribution of X in the control group. To formalize this phenomenon, we introduce notation for the conditional expectations in Table 5, where group is indicated by j (Control:j=0, Index:j=1).

Table 5

Strata Definitions and Conditional Expectations

Stratum	j	Interval	Label	$E(Y\|j,X\in A_k)$
Control	0	$(-\infty, c_{10})$	A_0	μ_{y0}
Low	1	$(-\infty, c_{11})$	A_1	μ_{y1}
Medium	1	(c_{11},c_{21})	A_2	μ_{y2}
High	1	(c_{21},∞)	A_3	μ_{y3}

The check mark pattern is said to hold if $\mu_{y0}>\mu_{y1}$ and $\mu_{y1}<\mu_{y2}<\mu_{y3}$. In the AFHS, c_{10}=10 ppt, c_{11}=10 ppt and c_{21}=24 ppt and in the CSF-EAA study $c_{10}=\infty$, c_{11}=31 and c_{21}=40.

3.2 Statistical Models to Predict the Check mark Pattern

We propose standard linear statistical models to predict the observed patterns in Tables 2 and 4. Let $\delta_{ji} = 1$ if subject i is in group j, 0 otherwise, j=0,1, i=1, 2, ..., n_j, and consider the following linear model for subject i in group j,

$$Y_{ji} = \beta_0 + \beta_1[(X_{0i} - \mu_{x0})\delta_{0i} + (X_{1i} - \mu_{x1})\delta_{1i}] + \varepsilon_{ji}, \tag{1}$$

where, for subject i in group j, j=0,1, i=1,2, ..., n_j, Y_{ji} and X_{ji} are the observed values of Y and X, μ_{xj} is the population mean of X from group j and ε_{ji} is a random error independent of X_{ji} and δ_{ji} with mean 0 and variance σ^2. Under model (1), $E(Y_{0i})=E(Y_{1i})=\beta_0$, that is, the overall means on Y in the two groups are equal. Furthermore, if $\beta_1>0$, it is easy to verify that model (1) predicts the check mark pattern. Using model (1), we have

$$\mu_{y0}=\beta_0+\beta_1[E(X_{0i}|X_{0i}\leq c_{10})-E(X_{0i})]. \tag{2}$$

But, since c_{10} is a high quantile of the control distribution of X, the coefficient of β_1 in (2) is approximately zero. However,

$$\mu_{y1}=\beta_0+\beta_1\,[E(X_{1i}|X_{1i}\leq c_{11})-E(X_{1i})] \tag{3}$$

and the coefficient of β_1 in (3) is negative because c_{11} is an approximate lower tertile of the distribution of X in the index group. Thus, $\mu_{y0}>\mu_{y1}$. The remaining conditional means follow the ordering $\mu_{y1}<\mu_{y2}<\mu_{y3}$ because $\beta_1>0$.

In the special case that all of the controls have X=0 and the index subjects have positive values of X, as in the CSF-EAA study, we reduce model (1) to

$$Y_{ji} = \beta_0 + \beta_1[(X_{i1} - \mu_{x1})\delta_{1i}] + \varepsilon_{ji}. \tag{4}$$

Once again, it is easy to verify that the reduced model (4) predicts equal means and the check mark pattern when $\beta_1>0$.

Model (1) implies that β_0 is the population mean of group 1 as well as of group 0. Therefore, β_0 is estimated by $\overline{Y}_0$, $\overline{Y}_1$ or $\overline{Y}=(n_0\overline{Y}_0+n_1\overline{Y}_1)/(n_0+n_1)$. Assuming a common variance in the two groups, it is better to estimate β_0 with $\overline{Y}$ because its variance is smaller than that of $\overline{Y}_0$ or $\overline{Y}_1$. Therefore, to accommodate this in our models, so that standard statistical software packages can be used, at the time of fitting the models, we replace the means μ_{xj} by their estimators $\overline{X}_j$, j=0,1. Based on (1), the conditional log(insulin) means are $\hat{\mu}_{y0}=4.24$, $\hat{\mu}_{y1}=4.08$, $\hat{\mu}_{y2}=4.26$ and $\hat{\mu}_{y3}=4.49$, which are very comparable to the observed means in Table 2. The predicted conditional means using model (4) for taurine, are $\hat{\mu}_{y0}=6.25$, $\hat{\mu}_{y1}=4.68$, $\hat{\mu}_{y2}=6.37$ and $\hat{\mu}_{y3}=8.03$, corresponding to the observed means in Table 4. Thus, model (1) and its reduction (4) describe the check mark patterns given in the two examples.

3.3 Reverse Causation

At this point we should mention that Flanders et al (1992) conducted a study of the check mark pattern using AFHS data. They studied two models; their first model considered a relationship between health and dioxin with health as the dependent variable and dioxin as the independent variable and their second model reversed the roles of these two variables. When analyzing their second model, they made the assumption that $E(Y|j=0)=E(Y|j=1)$, i.e. the Ranch Hand and Control health means are equal, while not making this assumption for their first model. As we show in the Appendix, under this

assumption the coefficient of health in their second model is identically zero (which was not recognized by Flanders et al), making their model not very useful. This mathematical oversight caused them to believe that their second "reverse causation" model was the appropriate model and hence concluded that the check mark pattern suggested reverse causation. Therefore, the mathematical arguments used by Flanders et al to argue that the check mark pattern suggests reverse causation are incorrect.

3.4 Estimating Exposure Effects

Because the concept of "exposure" is central to the AFHS, one may define a subject as exposed (e=1) if his dioxin level is above c_{10} and unexposed (e=0) otherwise. Using model (1), the mean of Y among the exposed, μ_{yE}, can be written as

$$\begin{aligned}\mu_{yE} &= E(Y_{ji}|e=1,\ j=0)P(j=0|e=1)+E(Y_{ji}|e=1,\ j=1)P(j=1|e=1)\\ &= \{\beta_0 \ +\beta_1[E(X_{0i}|X_{0i}>c_{10})-\mu_{x0}]\}\ \frac{P(e=1|j=0)P(j=0)}{P(e=1)}\\ &\quad +\{\beta_0 \ +\beta_1[E(X_{1i}|X_{1i}>c_{10})-\mu_{x1}]\}\ \frac{P(e=1|j=1)P(j=1)}{P(e=1)}.\end{aligned}$$

Similarly, the mean of Y among the unexposed is

$$\begin{aligned}\mu_{y\bar{E}} &= E(Y_{ji}|e=0,\ j=0)P(j=0|e=0)+E(Y_{ji}|e=0,\ j=1)P(j=1|e=0)\\ &= \{\beta_0 \ +\beta_1[E(X_{0i}|X_{0i}\le c_{10})-\mu_{x0}]\}\ \frac{P(e=0|j=0)P(j=0)}{P(e=0)}\\ &\quad +\{\beta_0 \ +\beta_1[E(X_{1i}|X_{1i}\le c_{10})-\mu_{x1}]\}\ \frac{P(e=0|j=1)P(j=1)}{P(e=0)}.\end{aligned}$$

However, the probabilities P(j=0) and P(j=1) are generally not estimable. Hence we consider the estimable difference of the Y mean of exposed index subjects and the Y mean

of unexposed controls, given by $D = \mu_{yE1} - \mu_{y\bar{E}0}$, where $\mu_{yE1} = E(Y|e=1, j=1)$ and $\mu_{y\bar{E}0} = E(Y|e=0, j=0)$ which, under model (1), is given by

$$D = \beta_1[p_{\bar{E}1}(\mu_{xE1} - \mu_{x\bar{E}1}) + p_{E0}(\mu_{xE0} - \mu_{x\bar{E}0})],$$

where p_{Ek}=P(e=1|j=k), $p_{\bar{E}k} = P(e=0|j=k)$, $\mu_{xEk} = E(X|e=1, j=k)$ and $\mu_{x\bar{E}k} = E(X|e=0, j=k)$, k=0,1. Hence D is estimated by

$$\hat{D} = \hat{\beta}_1[\hat{p}_{\bar{E}1}(\overline{X}_{E1} - \overline{X}_{\bar{E}1}) + \hat{p}_{E0}(\overline{X}_{E0} - \overline{X}_{\bar{E}0})],$$

where $\overline{X}_{Ek}$ and $\overline{X}_{\bar{E}k}$ are the sample means of exposed and unexposed subjects in group k, k=0,1, $\hat{p}_{E1}$ is the sample proportion of exposed subjects in group 1 and $\hat{p}_{\bar{E}0}$ is the sample proportion of unexposed subjects in group 0. The estimate of the conditional standard deviation of $\hat{D}$, given X, is

$$\hat{\sigma}_{\hat{D}} = \hat{\sigma}_{\hat{\beta}_1}[\hat{p}_{\bar{E}1}(\overline{X}_{E1} - \overline{X}_{\bar{E}1}) + \hat{p}_{E0}(\overline{X}_{E0} - \overline{X}_{\bar{E}0})].$$

Hence, to test the hypothesis H_o:D=0, we use the statistic $T = \hat{\beta}_1 / \hat{\sigma}_{\hat{\beta}_1}$, distributed as t with n-2 degrees of freedom under H_o. In the special case that all of the index subjects are exposed and all of the controls are unexposed, D reduces to the difference of group means E(Y|j=1)-E(Y|j=0) and the appropriate test statistic becomes the ordinary two sample t-test. For example, to estimate D in the Air Force Health Study data on log(insulin) and dioxin, we have $\hat{\beta}_1 = 0.18$, $\hat{\sigma}_{\hat{\beta}_1} = 0.031$, $p_{\bar{E}1} = 0.44$, $p_{E0} = 0.02$, $\overline{X}_{E1} = 3.384$, $\overline{X}_{\bar{E}1} = 1.836$, $\overline{X}_{E0} = 2.705$ and $\overline{X}_{\bar{E}0} = 1.501$. Therefore, $\hat{D} = 0.13$ and $\hat{\sigma}_{\hat{D}} = 0.022$.

Hence a 95% confidence interval for D is 0.09 to 0.17 and the test statistic for H_o is $T=5.81$ ($p<0.001$).

4. REFERENCES

Anderson, S., Auquier, A., Hauck, W.W., Oakes, D., Vandele, W., Weisberg, H.I., Bryk, A.S. and Kleinman, J. (1980). Statistical methods for comparative studies. John Wiley and Sons, New York.

Flanders, W.D., Lin, L., Pirkle, J.L., and Caudill, S.P. (1992), "Assessing the direction of causality in cross-sectional studies," The American Journal of Epidemiology, 135, 926-935.

Giovengo, S., Larsen, A.A., Michalek, J.E. and Orr, M.D. (1995). Essential amino acids fibromyalgia syndrome cerebral spinal fluid. J Musculoskeletal Pain 3 (supple #1) (to appear)

Grubbs, W.D., Lustik, M.B., Brockman, A.M., Henderson, S.C., Burnett, F.R., Land, R.G., Osborne, D.J., Rocconi, V.K., Schreiber, M.E., and Williams, D.E. (1995), The Air Force Health Study. An Epidemiologic Investigation of Health Effects in Air Force Personnel Following Exposure to Herbicides. 1992 Follow-up Examination Results, Armstrong Laboratory: Brooks Air Force Base, Texas.

Institute of Medicine (1994), Veterans and Agent Orange. Health Effects of Herbicides Used in Vietnam, Washington DC: National Academy Press.

Hornfeldt, C.S., Smullin, D.H., Schamber, C.D., Sun, X. and Larsen, A.A. (1992). Antinociceptive effects of intrathecal taurine and calcium in the mouse. Life Sciences 50, 1925-1934.

Larsen, A.A. (1989). Intrathecal GABA, glycine, taurine, or beta-alanine elicits dyskinetic movements in mice. Pharmacol Biochem Behav.

Malmberg, A., Yaksh, T. (1992). Hyperalgesia mediated by spinal glutamate or substance P receptor blocked spinal cyclooxygenase ihibition. Science 257, 1276-1279.

Michalek, J.E., Wolfe, W.H., and Miner, J.C. (1990), "Health status of Air Force veterans occupationally exposed to herbicides in Vietnam. II. Mortality," Journal of the American Medical Association, 264, 1832-1836.

Michalek, J.E., Pirkle, J.L., Caudill, S.P., and Tripathi,R.C., Patterson Jr, D.G., and Needham, L.L. (1995)., "Pharmacokinetics of TCDD in veterans of Operation Ranch Hand: 10-year follow-up," (To appear in the Journal of Toxicology and Environmental Health).

Palkovitz, M., Elekes, I., Lang, T. and Patthy, A. (1986). Taurine levels in discrete brain nuclei of rats J Neurochem 47, 1333-1335.

Russell, I.J., Vipraio, G.A., Morgan, W.W. and Bowden, C.L. (1986). Is there a metabolic basis for the fibrositis syndrome? Am J Med 81 (suppl 3A): 50-54.

Sharma, H.S., Nyberg, F., Olsson, Y. and Dey, P.K. (1990). Alterationof substance P after trauma to the spinal cord: an experimental study in the rat. Neuroscience 38, 205-212.

Skilling, S.R., Smullin, D.H. and Larsen A.A. (1990). Differential effects of C- and N-terminal substance P metabolites on the release of amino acid neurotransmitters from the spinal cord: potential role in nociception. Journal of Neuroscience 10, 1309-1318.

Smullin, D.H., Skilling, S.R. and Larsen, A.A. (1990). Interactions between substance P, calcitonin gene-related peptide, taurine and excitatory amino acids in the spinal cord. Pain 42, 93-101.

Wolfe, W.H., Michalek, J.E., Miner, J.C., Rahe, A., Silva, J., Thomas, W.F., Grubbs, W.D., Lustik, M.B., Karrison, T.G., Roegner, R.H., and Williams, D.E. (1990), "Health status of Air Force veterans occupationally exposed to herbicides in Vietnam. I. Physical health," Journal of the American Medical Association, 264, 1824-1831.

Wolfe, W.H., Michalek, J.E., Miner, J.C., Needham, L.L., and Patterson Jr, D.G. (1992), "Diabetes versus dioxin body burden in veterans of Operation Ranch Hand," in Organohalogen Compounds. Volume 10, Toxicology, Risk Assessment and Management, pp 279-282, Helsinki: Finnish Institute of Occupational Health.

6. APPENDIX

In our notation, Flanders et al Model 2 for subject i, i=1,2, ..., n, is given by

$$X_{2i}=a+bX_{1i}+cY_i+\varepsilon_i,$$

where X_2 is measured dioxin, X_1 is true dioxin, Y is the health variable and ε is error. The four variables, X_1, X_2, Y and ε are assumed jointly normally distributed with $E(X_1)=\mu_1$, $VAR(X_1)=\sigma_1^2$, $E(Y)=\mu_y$, $VAR(Y)=\sigma_y^2$, $E(\varepsilon_3)=0$ and $VAR(\varepsilon)=\sigma^2$ and X_1, Y and ε are assumed mutually independent. Thus health (Y) and measured dioxin X_2 are bivariate normal with mean vector $(\mu_y,\ a+b\mu_1+c\mu_y)$, $VAR(X_2)=c^2\sigma_y^2+b^2\sigma_1^2+\sigma_\varepsilon^2$, and $COV(Y,X_2)=c\sigma_y^2$. Therefore $E(Y|X_2=x_2)=\gamma_0+\gamma_1 x_2$, where

$$\gamma_0=\mu_y-\left\{\frac{c\sigma_y^2(a+b\mu_1+c\mu_y)}{c^2\sigma_y^2+b^2\sigma_1^2+\sigma_\varepsilon^2}\right\}$$

and

$$\gamma_1=\frac{c\sigma_y^2}{c^2\sigma_y^2+b^2\sigma_1^2+\sigma_\varepsilon^2}.$$

But Flanders et al also assume $E(Y|j=0)=E(Y|j=1)$, where j=0 for control subjects and j=1 for index subjects. which implies

$$\gamma_0+\gamma_1 E(X|j=0)=\gamma_0+\gamma_1 E(X|j=1).$$

But $E(X_2|j=1)>E(X_2|j=0)$. Hence, $\gamma_1\equiv 0$ and, therefore, $c\equiv 0$. Thus, Model 2 and their assumption that $E(Y|j=1)=E(Y|j=0)$ imply that Y and X_2 are unrelated. As a result, $k\equiv 0$ in their equation (8) and so their conclusions regarding the check mark pattern and reverse causation are not correct.

Figure 1

Log(Insulin) and Log(Dioxin+1) in the 1992 Air Force Health Study

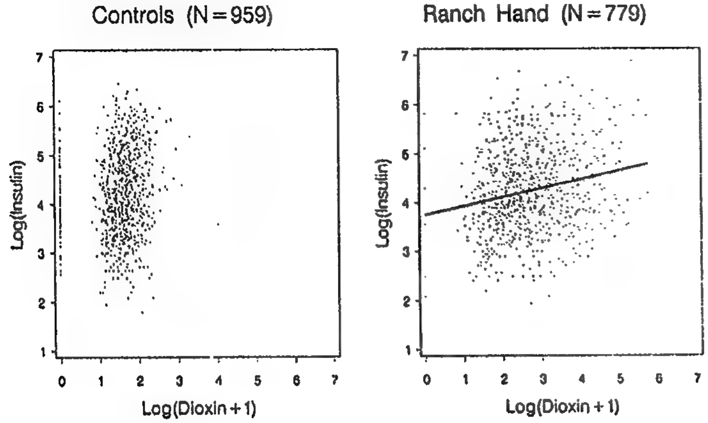
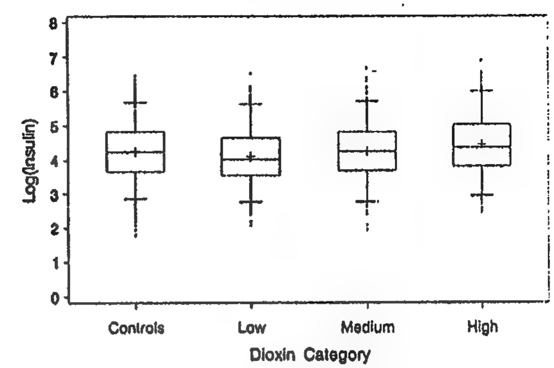


Figure 2

Log(Insulin) by Dioxin Category in the Air Force Health Study



LEGEND: Low, Medium and High are Dioxin categories in Ranch Hand patients. Whiskers extend to the 5th and 95th percentiles. Dots are lower and upper 5 percent. Rectangles are determined by quartiles and the mean is indicated with a +.

Figure 3

Taurine in CSF and the Tender Point Index in the Excitatory Amino Acid Study

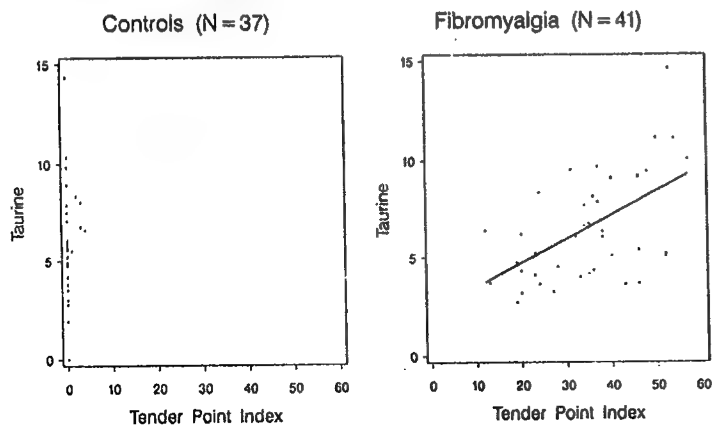
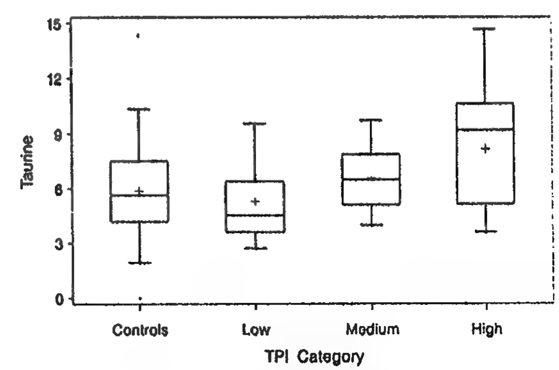


Figure 4

Taurine by TPI Category in the Excitatory Amino Acid Study



LEGEND: Low, Medium and High are TPI categories in FS patients. Whiskers extend to the 5th and 95th percentiles. Dots are lower and upper 5 percent. Rectangles are determined by quartiles and the mean is indicated with a +.

THE SUBSTRATE AND INDUCER SPECIFICITY OF THE NITROBENZENE NITROREDUCTASE PRODUCED BY *PSEUDOMONAS PSEUDOALCALIGENES* JS45

Michael P. Labare

Assistant Professor

Department of Chemistry

United States Military Academy

West Point, NY 10996

Final Report for:

Summer Faculty Research Program

Armstrong Laboratory

Sponsored by:

Air Force Office of Scientific Research

Bolling Air Force Base, DC

and

Armstrong Laboratory

August 1995

THE SUBSTRATE AND INDUCER SPECIFICITY OF THE NITROBENZENE NITROREDUCTASE PRODUCED BY *PSEUDOMONAS PSEUDOALCALIGENES* JS45

Michael P. Labare

Abstract

Pseudomonas pseudoalcaligenes JS45 uses nitrobenzene as the sole source of C, N, and energy via the incomplete reduction of the nitro group to hydroxylaminobenzene. However, JS45 does not grow on other nitroaromatics (Nishino and Spain, 1993). This study examined the substrate specificity and products formed by the nitrobenzene nitroreductase and the ability of various nitroaromatics to induce production of the nitroreductase. The nitroreductase reduced a wide range of nitroaromatics: 2-, 3-, and 4-nitrophenol; 2-, and 4-nitrobenzoate, 1,2-, 1,3-, and 1,4-dinitrobenzene, 2-, 3-, and 4-nitrotoluene, 2,3- 2,4-, 2,6-, and 3,4-dinitrotoluene, 2-, 3-, and 4-nitroaniline, 4-amino-3-nitrotoluene, trinitrobenzene, trinitrotoluene, nitroquinone, 4-nitrobenzyl alcohol, and 4-nitrobenzaldehyde. Of those tested only 2-nitrobenzoate and 4-nitrocatechol were not reduced to a significant extent. The experiments to determine the products formed from these substrates and their ability to induce production of the reductase are still in progress. The lack of growth of JS45 on nitroaromatics can not be attributed to a limited substrate range of the nitroreductase since the enzyme transforms a variety of nitroaromatics. The substrates may not induce the reductase or the products may be toxic or not substrates for subsequent enzymes in the degradative pathway.

THE SUBSTRATE AND INDUCER SPECIFICITY OF THE NITROBENZENE NITROREDUCTASE PRODUCED BY *PSEUDOMONAS PSEUDOALCALIGENES* JS45

Michael P. Labare

Introduction

Nitroaromatic compounds are produced in the synthesis of solvents, dyes, and explosives and are commonly found as environmental contaminants. Some (e.g. nitrobenzene and 2,4-dinitrotoluene) have been listed as priority pollutants by the Environmental Protection Agency (Keith and Telliard, 1979). Most nitroaromatic compounds are biodegradable and many can be completely mineralized by bacteria which utilize them as growth substrates (Gorontzy et al, 1994). The simplest nitroaromatic compound, nitrobenzene (NB), is biodegradable (Patil and Shinde, 1989) but until recently little was known about the degradative pathways.

Pseudomonas pseudoalcaligenes JS45 uses NB as a sole source of carbon, nitrogen, and energy. The degradative pathway starts with the reduction of NB to hydroxylaminobenzene (HAB), followed by conversion of HAB to 2-aminophenol (2AP), which is cleaved to form 2-aminomuconic semialdehyde (Nishino and Spain, 1993). The enzyme responsible for initial reduction is nitrobenzene nitroreductase. The nitroreductase has been purified and characterized. It is an oxygen-insensitive (Type 1) flavoprotein (2 moles of FMN per mole of protein) and is active as a monomer of 30-33 kDa. The reductase does not require metal cofactors and its production is induced by nitrobenzene (Somerville et al, 1995).

P. pseudoalcaligenes JS45 grows on NB, HAB, and 2AP but not on a variety of other aromatic compounds. Compounds tested include 1,2-, 1,3-, and 1,4-dinitrobenzene, 2-, 3-, and 4-nitrobenzoate, 2-, 3-, and 4-nitrotoluene, 2-, 3-, and 4-nitrophenol, catechol, 3- and 4-

nitrocatechol, 3- and 4-aminophenol, aniline, benzene, phenol, toluene, nitrobenzene, and picolinic acid (Nishino and Spain, 1993). Transformation of these compounds was not examined.

The lack of growth on these compounds may be due to a number of factors. The compounds may not serve as substrates for the reductase or the transformation products may not be substrates for the subsequent enzymes in the pathway. They may also not induce production of the reductase, mutase, or dioxygenase. In order to better understand the substrate range of *P. pseudoalcaligenes*, the substrate specificity and products formed by the nitroreductase were examined as well as the ability of various nitroaromatics to induce production of the nitroreductase.

Methodology

Substrate specificity. *P. pseudoalcaligenes* sp. strain JS45 nitroreductase was produced and purified as described in Somerville et al (1995). Reductase activity was measured spectrophotometrically as the decrease in A_{340} due to the oxidation of NADPH in the reaction mixture. Unless otherwise stated the reaction mixture contained enzyme, test compound (0.1 mM), NADPH (0.5 mM) and phosphate buffer (20 mM, pH 8) in 1 ml (Somerville et al, 1995). Test compounds were dissolved in water or ethanol. Specific activity was expressed as μmol NADPH oxidized per minute per mg of protein. All tests were run in triplicate.

Induction specificity studies. A series of preliminary experiments were conducted to determine the optimal conditions for the induction of nitrobenzene nitroreductase. JS45 was grown (30°C; 200 rpm) in 2 L of minimal medium (MSB; Stanier et al, 1966) amended with 20 mM succinate which received a 1% inoculum from an overnight culture grown in the same medium. When the 2-L culture reached an optical density (OD, 600 nm) of 0.5, the cells were harvested by centrifugation (10,000 g * 10 min.), washed two times with cold MSB, and

resuspended in 250-ml of MSB at an OD of 0.2. Replicate flasks containing either nitrobenzene, one test compound or no addition were incubated at 25°C, 240 rpm for 12 hours.

The test substrates were: nitrobenzene, 4-nitrotoluene, 4-nitrobenzoate, 4-nitrobenzaldehyde, 4-nitrobenzyl alcohol, 2-aminophenol, nitrosobenzene, 2-nitrophenol, 3-nitrophenol, 1,2-dinitrobenzene, 3-nitrobenzoate, 2,4-dinitrotoluene, 2,6-dinitrotoluene, hydroxylaminobenzoate, 2,4,6-trinitrotoluene. Test substrates were added at 1 mM save 4-nitrobenzaldehyde which was at 0.5 mM due to solubility problems; controls were unamended. The cells were then harvested, washed two times with cold phosphate buffer (20 mM, pH 7.5), and resuspended in 1-3 ml of cold buffer. The cells were broken by two passages through a French Press at 20,000 psi. The crude lysate was centrifuged at 14,000 g * 15 minutes and the pellet discarded.

Induction of all three enzymes was examined. Reductase activity was measured as described above with NB as the substrate. The mutase was assayed by spectrophotometrically following the production of 2AP from HAB at A_{282}. Reaction mixtures contained enzyme, HAB (0.5 mM), and tris buffer (50 mM, pH 7.0) in 1 ml. The dioxygenase was measured by monitoring the consumption of O_2 in reaction mixtures containing enzyme, 2AP (0.14 mM), and phosphate buffer (20 mM, pH 7.5) in 1.8 ml.

The toxicity of various nitroaromatics on JS45 was examined. JS45 was grown on MSB + 20 mM succinate as described above. The resuspended cells (OD = 0.01) were amended with nothing, 1.0 mM NB, or 1.0 mM NB and 1.0, 0.5, or 0.25 mM of a test compound. The OD was monitored for several days to determine if growth occurred.

Degradation of mixed substrates. JS45 in MSB amended with 1.0 mM nitrobenzene and 1.0 mM 2-, 3-, or 4-nitrophenol was grown as described above. Samples (1 ml) were removed at 0 and 24 hours. The loss of nitrobenzene and the nitrophenols was determined by HPLC analysis.

HPLC analysis. Samples (50 µl) of the reaction mixture described above were analyzed on a 250 mm Spherisorb C8 column (Alltech, Deerfield. IL) with a gradient of acetonitrile and trifluoroacetic acid (0.1 % in distilled water) as mobile phase. Elution started with 60% water and 40% acetonitrile and changed to 40 % water and 60% acetonitrile over 8 min. The flow rate was 1.0 ml/min. Products were detected at UV A_{230} with a HP1040A diode array detector (Hewlett-Packard Co., Palo Alto, CA). Products were identified by comparison of retention times and spectra with authentic standards.

Results.

Substrate specificity. The nitroreductase reduced a wide range of nitroaromatics (Table 1). Only, 2-nitrobenzoate and 4-nitrocatechol were not reduced. The nitroreductase also reduced benzoquinone with a rate of 20.2 µm of NADPH per min per mg protein (85% of the specific activity of NB) but had no activity on oxidized glutathione, tyrosine, phenylalanine, or cystine (Table 1).

Transformation products. Two products with retention times of 3.5 and 7.2 minutes were detected with 2-nitrophenol as the substrate. Based on a comparison of retention times and UV spectra, the product at 3.5 minutes has been tentatively identified as 2-aminophenol. The product at 7.2 minutes has not been identified. 3- and 4-nitrophenol also appear to produce the corresponding amines and at least one additional product. The other products have not been identified. At least one product was seen in the HPLC chromatograms of the nitroaromatics reduced by the nitroreductase. However, these products have not been identified to date.

Enzyme induction studies. The toxicity of several nitroaromatic compounds on JS45 was examined (Table 2). At 1.0, 0.5, and 0.25 mM, 3- and 4-nitrophenol, 2-, 3-, 4-nitrotoluene, 1,2-, 1,3-, and 1,4-dinitrobenzene, 4-nitrobenzaldehyde, 4-nitrobenzyl alcohol and 2,3-dinitrotoluene inhibited growth of JS45 on 1.0 mM NB when compared to its growth on 1.0 mM nitrobenzene alone. However, the OD of these cultures cells did not decrease below the OD of unamended cultures, i.e. starved cells. This indicated the compounds were not killing the cells and some other mechanism was responsible for the lack of growth. For this reason all the test substrates except 4-nitrobenzaldehyde were maintained at 1.0 mM. Several of the compounds (2-nitrophenol, 3-nitrobenzoate, and 2,4-dinitrotoluene) appeared to support greater growth than nitrobenzene (data not shown).

A preliminary experiment with test substrates (nitrobenzene, 2-, 3-, or 4-nitrophenol, 2-, 3-, or 4-nitrobenzoate, 1,2, 1,3-, or 1,4-dinitrobenzene) added to 250 ml of JS45 (OD = 1) at 0.1 mM was conducted. After 12 hours exposure to the nitroaromatics there was no statistical difference between the average specific activity of the reductase, mutase or dioxygenase in any of the treatments (data not shown). Additional experiments determined the induction conditions to be 25°C, 240 rpm, 250 ml, OD of 0.2, and 1 mM nitrobenzene. Under these conditions induction of the nitroreductase occurred 4 h prior to growth which started at 10 hours. However, this pattern was not repeatable. In subsequent experiments induction was examined after growth was seen with the nitrobenzene amended cells.

In the final induction experiments, growth of JS45 started after 15.5 hours. Reductase and mutase activities were assayed as described above. The specific activity for reductase from nitrobenzene grown cells was 20 μmol NADPH oxidized/minute/mg of protein (Table 3). Analysis of variance with the Tukey's honestly significant difference method (Sokal and Rohlf,

1981) indicated there was no statistical difference between the specific activities of any of the treatments.

The specific activity of the mutase for nitrobenzene-grown cells was 8.4 µmol 2-AP formed/minute/mg protein (Table 3). Oxygenase activities were not determined for this experiment.

Degradation of mixed substrates. After 24 hours, JS45 had completely degraded 1.0 mM nitrobenzene. The concentration of 2-, 3, or 4-nitrophenol decreased by 100, 32, and 25%, respectively. After 3 days the culture with nitrobenzene and 2-nitrophenol were reamended. Within 24 hours, only trace amounts of the nitroaromatics were detected. The addition of fresh nitrobenzene did not cause an appreciable increase in the amount of 3- or 4-nitrophenol degraded.

Discussion

JS45's inability to grow on most nitroaromatics could be due to several reasons. 1) The nitroaromatics may be toxic to JS45; 2) the compounds may not be substrates for the reductase; 3) they may not induce production of the reductase; or 4) the nitroaromatics which are transformed by the reductase may form products that are toxic or are not substrates for subsequent enzymes in the pathway.

The toxicity of nitroaromatics has been demonstrated (Won et al, 1976). In fact nitrobenzene becomes inhibitory to JS45 at concentrations exceeding 2 mM (Urs Lendenmann, personal communication). However, in this study it is not likely that the cells were being killed by the nitroaromatics since the OD of the cultures did not decrease over a 2-3 day period. It is possible to interpret these results as a lack of growth rather then toxicity i.e. metabolism either did not occur or stopped before C, N, or energy could be obtained.

The lack of growth can not be attributed to a limited substrate range of the nitroreductase. The reductase reduced a wide range of compounds, many with equal or greater specific activities as nitrobenzene e.g. 1,3-dinitrobenzene, 4-nitrotoluene, and 2,6-dinitrotoluene (Table 1). Thus it appears either the reductase was not induced or the transformation products were not substrates for the mutase or dioxygenase.

A lack of induction of the reductase is likely one reason for the lack of growth on other nitroaromatics. This is indicated by the experiments where the nitrophenols were simultaneously degraded along with nitrobenzene. However, the experiments to demonstrate induction of the reductase were not repeatable. The preliminary studies had shown induction by nitrobenzene occurred 4 hours prior to growth. However, there was no statistical difference between the specific activities of cells grown on nitrobenzene, starved, or exposed to another nitroaromatic (Table 3). The only difference between the two series of experiments was the history of the cells exposed to the nitroaromatics. In the preliminary studies the cells were grown in MSB amended with succinate until they reached an OD of approximately 0.5. At this OD the cells are in mid-exponential growth. In the later studies the cells had exceeded an OD of 0.5 and were diluted to an OD of approximately 0.4 with fresh medium and allowed to grow back to an OD of 0.5. It is possible that a majority of the cells were not in the same growth phase but rather were actually late exponential or stationary phase and unable to induce as quickly. Experiments are being planned to test this hypothesis.

There was no statistical difference between the mutase specific activities for the nitrobenzene and any of the other treatments, i.e. starved cells or exposed to a nitroaromatic (Table 3).

Additional studies are being conducted to isolate and identify the transformation products formed by the nitroreductase on various nitroaromatic compounds. The products will be tested as possible substrates for the mutase and dioxygenases. In addition, induction of the nitroreductase by the nitroaromatics will continued to be examined.

Acknowledgments

I would like to thank Dr. Jim Spain for the opportunity to work on this project in his laboratory and for his guidance. I would also thank Dr. Charles Somerville for his large contribution to this study, both in the planning and execution.

Table 1. Nitrobenzene reductase activity on nitroaromatic compounds

Day[a]	Substrate	Specific activity[b] (μmol of NADPH min^{-1} mg^{-1})	Percent of nitrobenzene activity
1	Nitrobenzene	81 ± 7.3	100
	2-Nitrophenol	9.9 ± 0.2	12
	3-Nitrophenol	5.5 ± 0.2	6.7
	4-Nitrophenol	5.1 ± 0.2	6.3
2	Nitrobenzene	77 ± 7.9	100
	2-Nitrobenzoate	0.3 ± 0.0	0.3
	3-Nitrobenzoate	3.7 ± 0.5	4.9
	4-Nitrobenzoate	25 ± 1.6	33
3	Nitrobenzene	47 ± 2.1	100
	1,2-Dinitrobenzene	44 ± 6.4	93
	1,3- Dinitrobenzene	62 ± 6.0	133
	1,4- Dinitrobenzene	15 ± 2.8	32
	2-Nitrotoluene	24 ± 1.2	52
	3-Nitrotoluene	46 ± 1.7	98
4	Nitrobenzene	43 ± 7.7	100
	4-Nitrotoluene	35 ± 8.3	81
	2,4-Dinitrotoluene	67 ± 6.1	155
	2,6- Dinitrotoluene	40 ± 0.4	94
	3,4- Dinitrotoluene	38 ± 1.6	87
	2,3- Dinitrotoluene	45 ± 2.7	104
5	Nitrobenzene	34 ± 3.0	100
	2-Nitroaniline	3.2 ± 0.3	9.5
	3-Nitroaniline	24 ± 2.3	71
	4-Nitroaniline	2.1 ± 0.02	6.4
	4-Amino-3-nitrotoluene	2.3 ± 0.3	7.0
	4-Nitrocatechol	0.06 ± 0.2	0
6	Nitrobenzene	31 ± 7.0	100
	Trinitrotoluene	18 ± 0.95	57
	Trinitrobenzene	20 ± 3.6	63
7	Nitrobenzene	24 ± 3.3	100
	Nitroquinone	20 ± 0.69	85
8	Nitrobenzene	24 ± 3.3	100
	4-Nitrobenzyl alcohol	14 ± 1.1	60
	4-Nitrobenzaldehyde	8.0 ± 0.6	33

[a]Reductase assays for the various compounds were run on different days with different batch of enzyme. The percent activity of the test compounds was determined using the specific activity of the enzyme for nitrobenzene run on that day.

[b] Mean ± standard deviation.

Table 2. Effects of nitroaromatic compounds on the growth of JS45 in MSB broth amended with 1.0 mM nitrobenzene

Test compound	1.0 (mM)	0.50 (mM)	0.25 (mM)
2-Nitrophenol	No effect	ND	ND
3-Nitrophenol	No growth	No growth	No growth
4-Nitrophenol	No growth	No growth	No growth
2-Nitrotoluene	Decreased growth	Decreased growth	Decreased growth
3-Nitrotoluene	No growth	Decreased growth	Decreased growth
4-Nitrotoluene	No growth	No growth	No growth
1,2-Dinitrobenzene	No growth	No growth	No growth
1,3-Dinitrobenzene	Increased lag	No growth	No growth
1,4-Dinitrobenzene	No growth	No growth	No growth
2-Nitrobenzoate	Decreased growth	ND	ND
3-Nitrobenzoate	Decreased growth	ND	ND
4-Nitrobenzoate	Decreased growth	ND	ND
4-Nitrobenzaldehyde	No growth	Decreased growth	No effect
4-Nitrobenzyl alcohol	No growth	No growth	No effect
2,3-Dinitrotoluene	Decreased growth	No effect	No effect
2,4- Dinitrotoluene	Decreased growth	ND	ND
2,6- Dinitrotoluene	Decreased growth	ND	ND
3,4- Dinitrotoluene	Decreased growth	ND	ND

ND = Not Done

Table 3. Specific activities of the nitroreductase and mutase of JS45 exposed to nitroaromatics for 15.5 hours

	Reductase specific activity[a] (μmol of NADPH min^{-1} mg^{-1})	Mutase specific activity[a] (μmol of 2-aminophenol formed min^{-1} mg^{-1})
Control	15 ± 4.4	8.4 ± 0.54
Nitrobenzene	20 ± 5.4	8.9 ± 1.6
4-Nitrotoluene	18 ± 4.9	8.8 ± 1.3
4-Nitrobenzoate	14 ± 5.7	11 ± 0.30
4-Nitrobenzaldehyde	20 ± 9.3	11 ± 1.7
4-Nitrobenzyl alcohol	19 ± 1.5	11 ± 0.88
2-Aminophenol	27 ± 1.2	9.7 ± 1.4
Nitrosobenzene	13 ± 8.7	13 ± 0.83
2-Nitrophenol	12 ± 7.2	13 ± 1.5
3-Nitrophenol	16 ± 5.9	13 ± 1.5
1,2-Dinitrobenzene	17 ± 7.0	9.7 ± 1.9
3-Nitrobenzoate	19 ± 7.9	9.2 ± 0.78
2,4-Dinitrotoluene	14 ± 1.0	9.9 ± 0.29
2,6-Dinitrotoluene	13 ± 8.6	8.0 ± 0.33
Hydroxyaminobenzene	10 ± 1.0	6.5 ± 0.28
2-Nitrobenzoate	12 ± 7.2	9.4 ± 1.4
Trinitrotoluene	13 ± 5.0	7.4 ± 1.0

[a]Mean ± standard deviation. There was no significant difference between the specific of any treatment and cells grown on nitrobenzene.($p \leq 0.05$).

References

Anlezark, G.M., R.G. Melton, R.F. Sherwood, B. Coles, F. Friedlos, and R. Knox. 1992. The bioactivation of 5-(aziridin-1-YL)-2,4-dinitrobenzamide (CB1954)-1. Purification and properties of a nitroreductase enzyme from *Escherichia coli* - a potential enzyme for antibody-directed enzyme prodrug therapy (ADEPT). Biochem. Parmacol. **44:**2289-2295.

Bryant, C. and M. DeLuca. 1991. Purification and characterization of an oxygen-insensitive NAD(P)H nitroreductase from *Enterobacter cloacae*. J. Biol. Chem. **266:**4119-4125.

Keith, L.Y. and W.A. and Telliard. 1979. Priority pollutants. I. A perpective view. Environ. Sci. Technol. **13:**416-423.

Groenewegen, P.E.J., P. Breeuwer, J.M.L.M. van Helvoort, A.A.M. Langenhoff, F.P. de Vries and J. A.M. de Bont. 1992. Novel degradative pathway of 4-nitrobenzoate in *Comamonas acidovorans* NBA-10. J. Gen. Microbiol. 138:1599-1605.

Gorontzy T., O. Drzyzga, M.W. Kahl, D. Bruns-nagel, J. Breitung, E. von Loew, and K.-H. Blotevogel. 1994. Microbial degradation of explosives and related compounds. Critical Reviews in Microbiol. **20:**265-284.

Nishino, S.F. and J.C. Spain. 1993. Degradation of nitrobenzene by a *Pseudomonas pseudoalcaligenes* Appl. Environ. Microbiol. **59:**2520-2525.

Patil, S.S. and V.M. Shinde. 1989. Gas chromatographic studies on the biodegradation of nitrobenzene and 2,4-dinitrophenol in the nitrobenzene plant wastewater. Environ. Pollut. **57:**235-250.

Sokal, R.R. and F.J. Rohlf. 1981. Biometry. W.H. Freeman and Co., N.Y.

Somerville, C.C., S.F. Nishino, and J.C. Spain. 1995. Isolation and characterization of nitrobenzene nitroreductase from *Pseudomonas pseudoalcaligenes* JS45. J. Bacteriol. **177:** 3837-3842.

Stanier, R.Y., N.J. Palleroni and M. Doudoroff. 1966. The aerobic pseudomonads: a taxonomic study. J. Gen. Microbiol. **43**:159-271.

Won, W.D., L.H. Disalvo and J. Ng. 1976. Toxicity and mutagenicity of 2,4,6-trinitrotoluene and its microbial metabolites. Appl. Environ. Microbiol. **31:** 576-580.

SOLID PHASE MICROEXTRACTION APPLIED TO THE PROBLEM OF FUEL SPILL IDENTIFICATION

Barry K. Lavine
Associate Professor
Department of Chemistry

Clarkson University
Box 5810
Potsdam, NY 13699-5810

Final Report for
Summer Faculty Research Program
Armstrong Laboratory

Sponsored by
Air Force Office of Scientific Research
Bolling Air Force Base, DC

and

Armstrong Laboratory

August 1995

SOLID PHASE MICROEXTRACTION APPLIED TO THE PROBLEM OF FUEL SPILL IDENTIFICATION

Barry K. Lavine
Department of Chemistry
Clarkson University

Abstract

Source identification is an important problem in the field of environmental chemistry. Pollutants must be related to their sources in order to develop effective strategies for the management of the environment. Using headspace/solid phase microextraction/capillary column gas chromatography as an alternative to conventional chromatographic methods, we show that jet fuels recovered from underground sampling wells can be identified as to source. This new approach to fuel spill identification possesses significant potential which can be exploited by the Air Force in its current fuel identification program.

SOLID PHASE MICROEXTRACTION APPLIED TO THE PROBLEM OF FUEL SPILL IDENTIFICATION

Barry K. Lavine

Introduction

Groundwater is the last remaining potable water resource for many households and communities in the United States (1). The contamination of this important natural resource by jet fuels stored in leaking underground tanks and pipelines is a serious environmental problem, which has prompted the United States Air Force to develop new methods for the identification of fuel materials recovered from the subsurface environment (2-3). The Air Force's interest in techniques that can establish the type of fuel responsible for a spill is motivated in part by the clean-up costs, legal fees and fines incurred by the polluter.

Water samples collected from underground wells contaminated by leaking fuels exist in one of two forms. Either water collected from the well shows layers of floating fuel or the water contains dissolved hydrocarbons from the leaking fuel. On-site chemical analysis of the fuel materials by solid phase microextraction (SPME) and gas chromatography (GC) is attractive because many of the problems (4-5) associated with collecting and transporting the water samples (e.g., representative sampling, contamination, loss of volatiles, storage, lack of temporal resolution, etc.) are eliminated. Furthermore, problems arising from transient environmental phenomena such as changes in the water table level or ambient temperature fluctuations which can significantly alter the apparent concentration of pollutants in groundwater as determined by standard laboratory techniques are mitigated by using in-situ analytical methods. Finally, on-site chemical analysis by SPME/GC possesses significant potential to circumvent the bottleneck caused by standard analytical methods such as purge and trap gas chromatography/mass spectrometry (6) or solid phase extraction (7).

In this report, we discuss headspace SPME/GC as a possible alternative to conventional chromatographic methods, such as purge and trap GC or solid phase extraction, for identification of fuel materials recovered from underground wells or aquifers. Although SPME has found much

application in direct extraction of semi-volatile organic compounds from water (8-9), the technique can also be applied to headspace and other vapor phase samples. For example, headspace SPME extractions have been reported for flavor components in food and beverages (10) and for environmental contaminants in water (11). The headspace technique is advantageous when the volatility of the analytes permits a headspace determination or when undesirable components in the bulk sample make direct SPME undesirable. Headspace sampling prolongs fiber life, and is faster than direct sampling (12). Finally, recovered fuels cannot be analyzed by direct SPME because of irreversible fiber damage cause by excessive swelling of the polydimethylsiloxane coating, which is a direct result of the fiber being in intimate contact with the fuel layer (13).

Solid Phase Microextraction

SPME is a microscale analytical extraction technique designed for use with gas chromatography. SPME employs a fused silica fiber coated with polydimethylsiloxane to extract organic compounds from aqueous samples. The fiber is attached to a modified syringe as shown in Figure 1. The organic analyte is extracted by introducing the syringe into the headspace of the sample for a period of time. When the plunger is depressed, the fiber is in contact with the sample, and organic compounds are absorbed by the polydimethylsiloxane coating. The extraction efficiency of the polydimethylsiloxane coating for an analyte is a function of the partition coefficient of the compound.

Following the extraction period, the plunger is retracted, and the SPME device is inserted directly into the injection port of the gas chromatograph. The plunger is then depressed, exposing the polymeric coating to the high temperature of the injector port, which ensures thermal desorption of the absorbed molecules. The absorbed molecules are immediately released into the carrier gas stream, where they are swept onto the column for separation and subsequent identification.

Experimental

Spill samples simulating ground water contaminated by a jet fuel were prepared by equilibrating a neat jet fuel with water in a vessel designed by Burris and McIntyre (14) to maximize

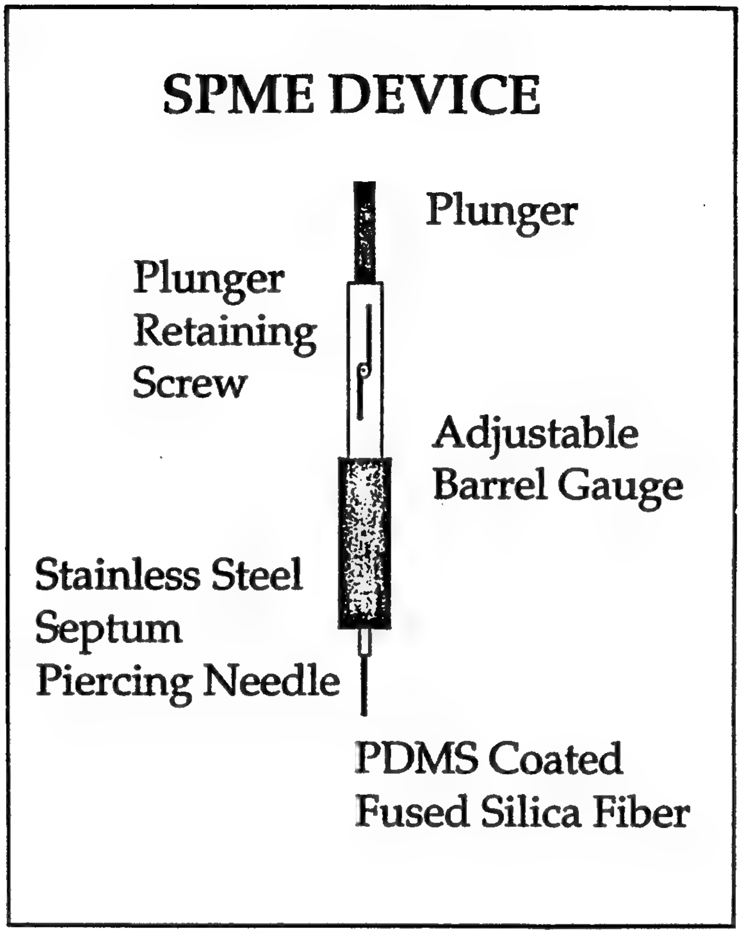


Figure 1. SPME Device.

surface contact between fuel and water while avoiding mixing (see Figure 2). 2 ml of a fuel were equilibrated with 250 ml of water for 12 hours while stirring gently. Following equilibration, several milliliters of water were discharged from the vessel to ensure the delivery tube was clear of fuel, and two 25-ml aliquots of the water phase were delivered into gas tight syringes equipped with Luer-lock open-shut valves. Hence, two 25-ml water samples can be prepared from a single fuel sample. In this study, 180 simulated water samples were prepared from 99 neat jet fuel samples (see Table 1).

Table 1. Water Soluble Data

Fuel Type	Number of Contaminated Water Samples	Number of Chromatograms
JP-4	20	36
Jet-A	29	50
JP-7	07	13
JPTS	10	20
JP-5	12	24
JP-8	21	37
Total	98	180

Each sample was placed in 40-ml VOA vials with one hole screw caps and teflon-faced septa. Prior to the introduction of the sample, a micro-stirring bar was placed in the vial to permit the sample to be stirred during the SPME sampling period which was 15 minutes. In a previous study (12), it was found that 15 minutes was sufficient to obtain a representative sampling of the water soluble compounds present in a jet fuel.

GC profiles of the microextracts were obtained using an HP-5890 gas chromatograph equipped with a flame ionization detector, a split/splitless injection port, and 30 meter fused silica capillary column (0.25mm in diameter, and coated with 1 micron of a bonded and cross-linked nonpolar stationary phase which consisted of 5% phenyl-substituted polymethylsiloxane (DB-5, J&W Scientific). The GC oven was temperature programmed from -10^{0}C, with an initial isothermal hold of 3 minutes to 250^{0}C, at a rate of 10^{0}C/minute, followed by a 6 minute final isothermal hold period.

SPME was also used in this study to sample the headspace of neat jet fuels. One hundred and seventy three fuel samples representing six different types of jet fuels (JP-4, Jet-A, JP-7, JPTS, JP-5,

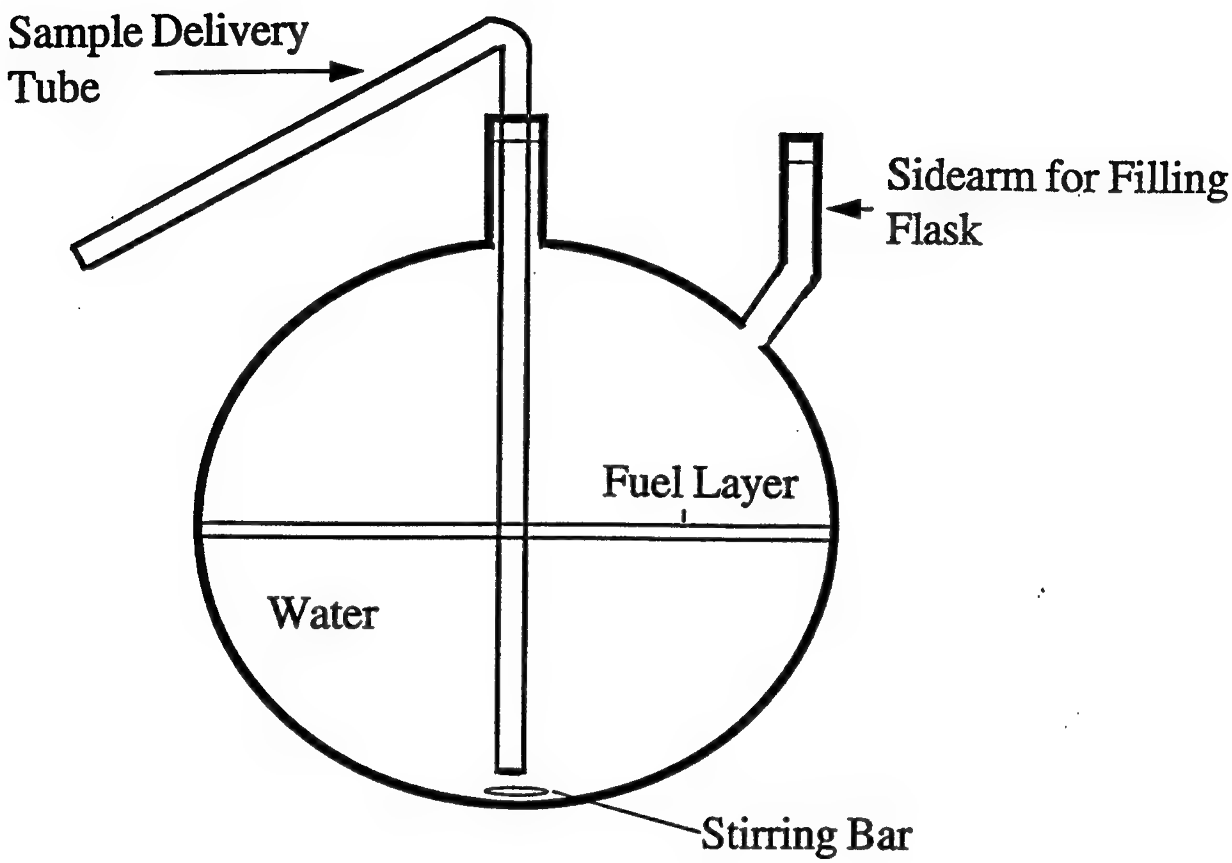


Figure 2. Flask for preparation of simulated water samples.

and JP-8) were analyzed by headspace SPME/GC (see Table 2). The fuel samples were obtained from Wright Patterson Air Force Base or Mukilteo WA Energy Management Lab and were splits from regular quality control standards used by these laboratories to verify the authenticity of manufacturer's claims. The control standards constituted a representative sampling of the fuels. Nineteen fuel samples (11 JP-4, 2 JP-5, 2 Diesel, 4 JP-8) recovered from the subsurface environment near Air Force and Navy airfields were also analyzed by headspace SPME/GC (see Table 2). The gas chromatograms of the recovered fuels were compared to the gas chromatograms of the neat jet fuels by visual analysis to seek a best match.

Table 2. Neat Jet Fuel Data

Fuel Type	Number of Samples
JP-4	46
Jet-A	45
JP-7	09
JPTS	15
JP-5	14
JP-8	44
Total	173

The following experimental protocol was used to obtain GC profiles of the volatile components of a jet fuel. Four milliliters of a neat jet fuel were placed in a 40 ml VOA bottle. A micro-stirring bar was also placed in the VOA bottle prior to the introduction of the fuel sample to permit stirring of the sample during the SPME sampling period which was 10 minutes. In a previous study, we found that 10 minutes was sufficient to obtain a representative sampling of the volatile compounds present in a jet fuel (12). Gas chromatograms of the microextracts were analyzed using a high efficiency fused silica capillary column 10 meters long with an internal diameter of 0.10 mm and coated with 0.34 µm of a bonded and cross-linked 5% phenyl-substituted polymethylsiloxane phase. The column was temperature programmed from 60 to 270^0 degrees Centigrade at 10^0 per minute.

Results & Discussion

Figure 3 shows GC profiles representative of water samples contaminated by neat jet fuels:

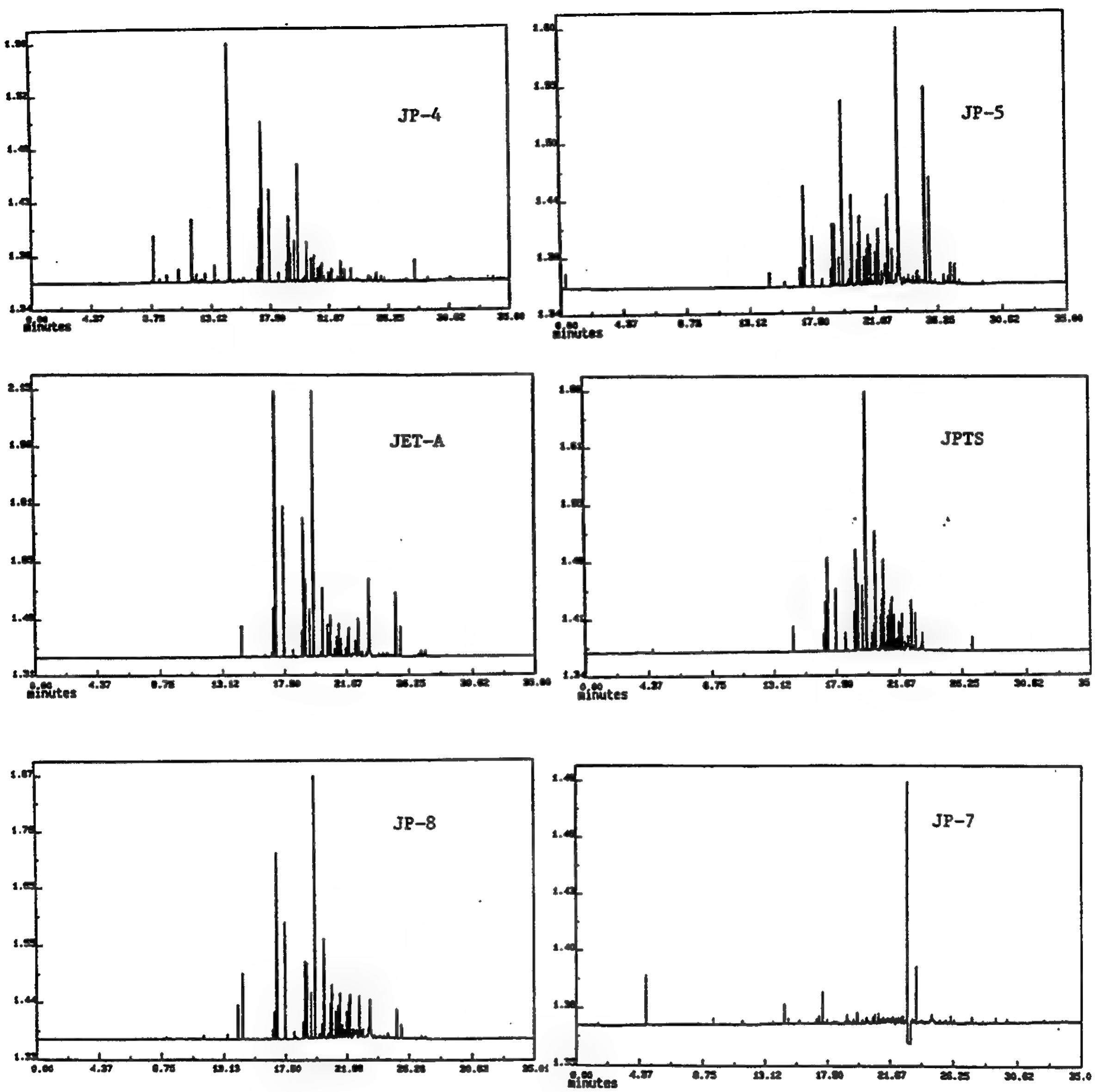


Figure 3. Headspace SPME/GC profiles of water samples contaminated by neat jet fuels.

the so-called simulated water samples. The GC profiles shown are representative of the prototypical vector of each fuel class. It is evident from an examination of the figure that each jet fuel possesses a characteristic profile, which suggests that information about fuel type can be obtained from gas chromatograms of hydrocarbons dissolved in water when SPME is used to sample the headspace of aqueous solutions contaminated by neat jet fuels. The characteristic nature of the GC data also suggests application of pattern recognition methods to identify fingerprint patterns in the data characteristic of fuel type.

Figure 4 shows GC profiles representative of the volatile components of neat jet fuels. The headspace of the neat jet fuels was sampled by SPME, and the gas chromatograms are representative of the prototypical vector of the six fuel classes shown. JP-4 and Jet-A fuels can be readily differentiated from the other fuels on the basis of their GC profiles. In fact, JP-4 fuel materials recovered from several monitoring wells at Tyndall could be identified as to fuel type by comparing the gas chromatograms of the recovered fuels to the gas chromatograms of the neat jet fuels, in order to seek a best match (see Figure 5). On the other hand, it is apparent from an examination of Figure 4 that GC profiles of JP-7, JPTS, JP-8 and JP-5 fuels share a similar set of attributes which suggests that it would be more difficult to reliably identify these fuels on the basis of their GC profiles.

The SPME experiments described in Figures 3 and 4 can be viewed as an example of selective fractionation of complex mixtures. The ease of classifying highly complex mixtures by this approach becomes apparent when considering the advantages of employing features from both sets of data to classify the jet fuels. JP-7 has very little aromatic content and hence few water soluble components, but contains significant amounts of straight and branched chain alkanes, alkenes, and alkyldienes, so classifying JP-7 fuels will be a simple task when features from both data sets are used to develop the discriminant. The volatile fraction of JP-4 provides a more characteristic profile of the fuel than GC's of the water soluble fraction, whereas the opposite holds true for JP-5, Jet-A, and JP-8 fuels which is an important result because of the strong similarities in gas chromatograms of the overall hydrocarbon profiles of these three fuels (15). In view of the fact that an equilibration time of only 3 hours is necessary to obtain a reproducible profile of the water soluble components of a jet fuel (see

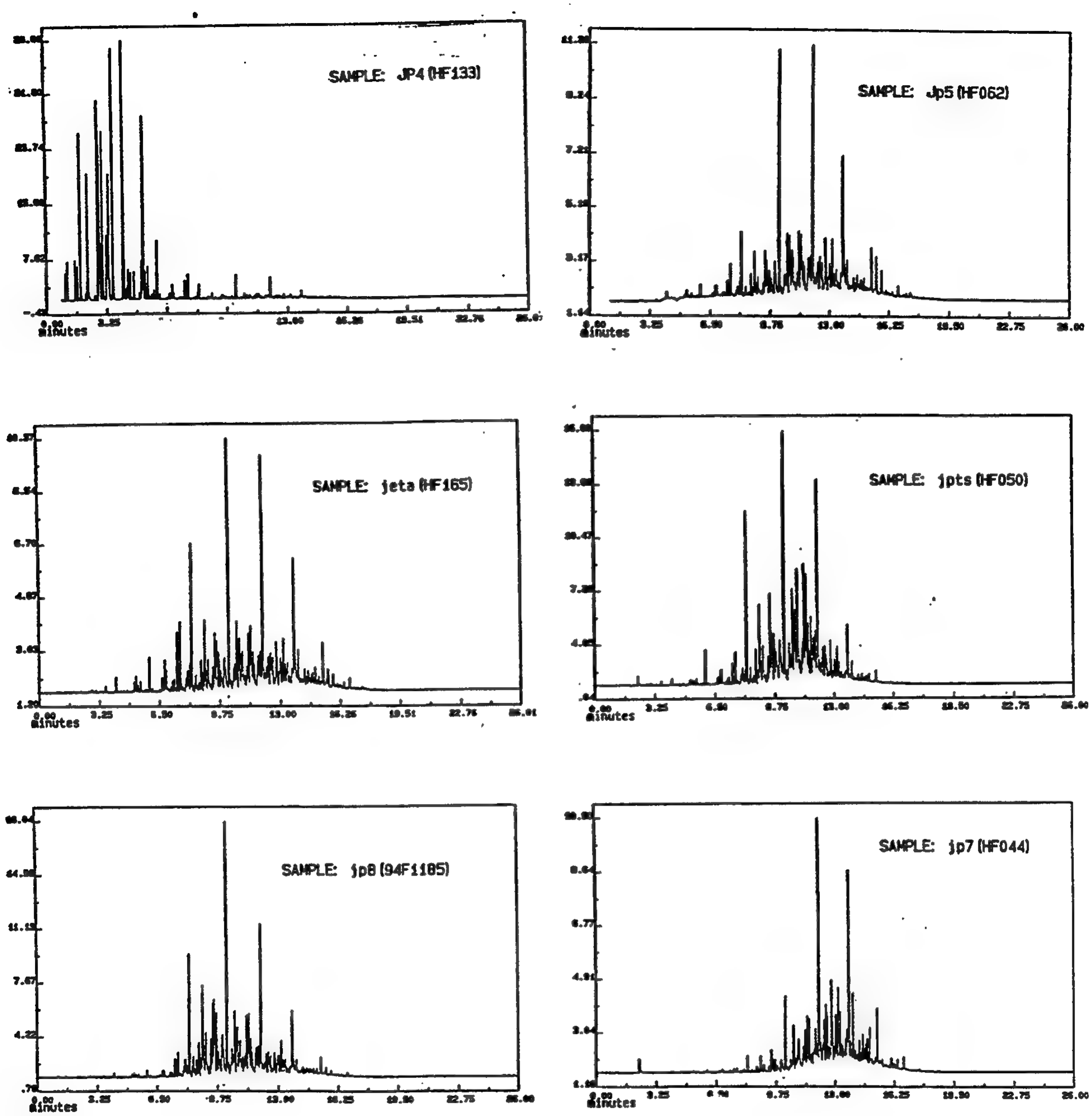


Figure 4. Headspace SPME/GC profiles of neat jet fuels.

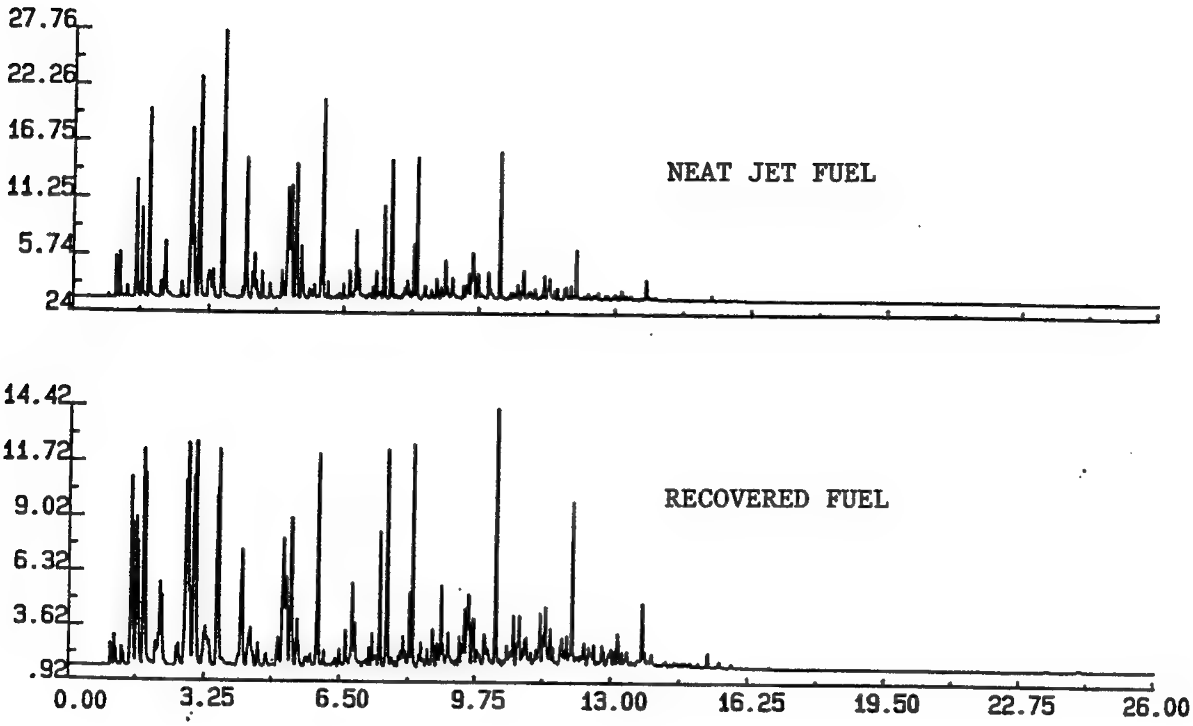


Figure 5. Comparison of the SPME/GC profile of a JP-4 fuel and a recovered fuel.

Figure 6), it seems reasonable to consider the direct implementation of the selective fractionation scheme described here as an integral component of the methodology used by the Air Force to identify recovered fuels from the environment.

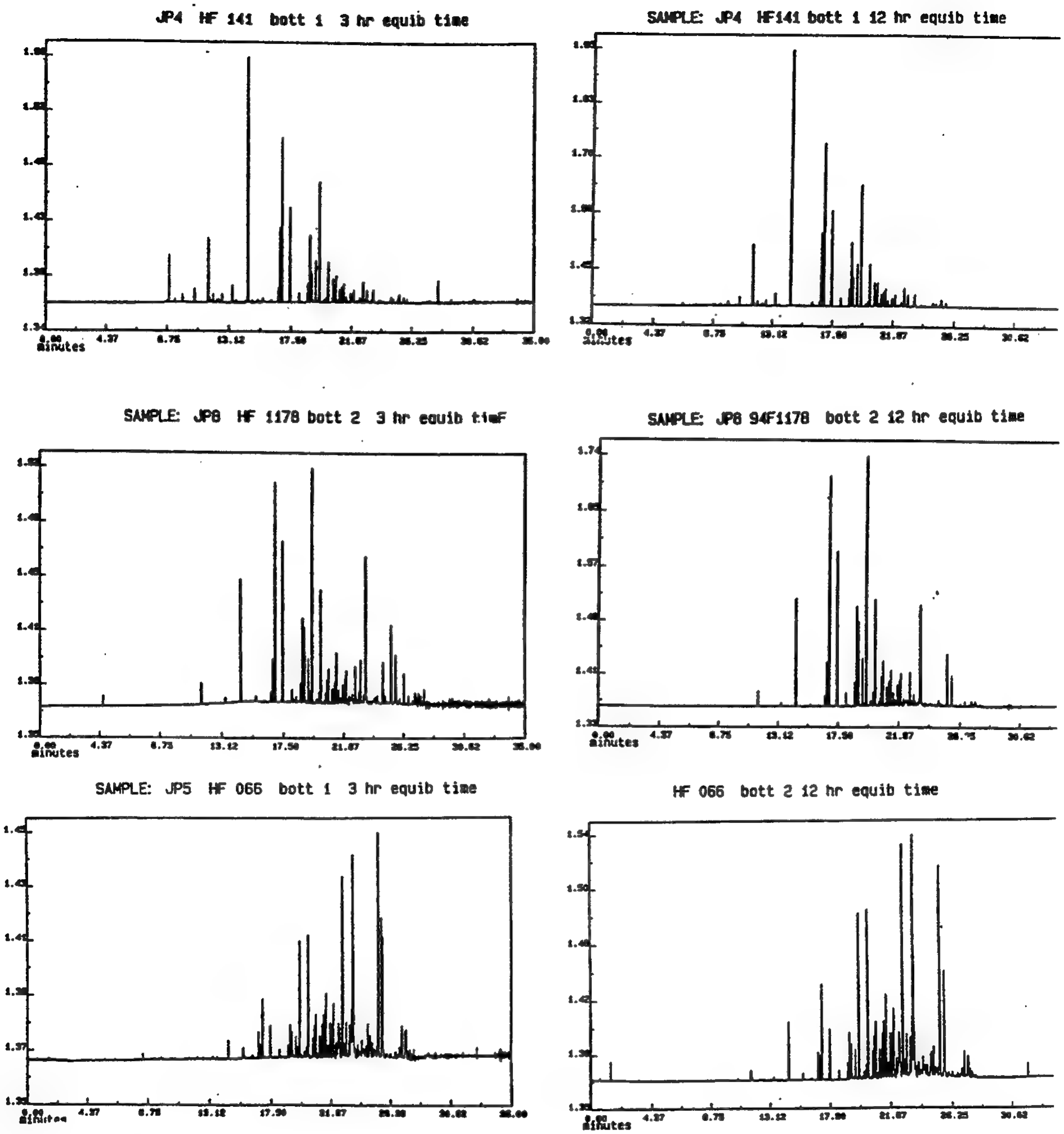


Figure 6. SPME/GC profiles of water soluble fraction as a function of equilibration time.

REFERENCES

1. E.M.Thurman, Organic Geochemistry of Natural Waters, Martinus Nijhoff/Dr W. Junk Publishers, Dordrecht, The Netherlands, 1985.

2. B.K.Lavine, A.Stine, and H.T.Mayfield, **Anal.Chim.Acta.**, 1993, 277, 357-367.

3. B.K.Lavine, X.Qin, A.Stine, and H.T.Mayfield, **Process Control & Quality**, 1992, 2, 347-355.

4. F.P.Milanovich, D.G.Garvis, S.M.Angel, S.M.Klainer, and L.Ecceles, **Anal. Instrum.**, 1986, 15, 137-147.

5. J.M.E.Storey, T.E.Barber, R.D.Shelton, E.A.Wachter, K.T.Carron, and Y.Jiang, **Spect.**, 1995, 10(3), 20-25.

6. R.L.Grob, Modern Practices in Gas Chromatography, John Wiley & Sons, New York, 1977, Chapter 3.

7. N.Simpson, Am.Lab. August 1992, p.37.; M.Zief and R.Kiser, **Am.Lab.**, January 1990, p.70.

8. C.L.Arthur, L.M.Killam, S.Motlahg, M.Lim, D.W.Potter, and J.Pawliszyn, **Environ.Sci.Techn.**, 1992, 26, 979-983.

9. C.L.Arthur, K.Pratt, S.Motlahg, J.Pawliszyn, and R.P.Belardi, **J.H.Res.Chrom.Comm.**, 1992, 15, 741-744.

10. S.B.Hawthorne, D.Miller, J.Pawliszyn, and C.L.Arthur, **J.Chrom.**, 1992, 603, 185-191.

11. C.Arthur, L.M.Killam, S.Motlagh, M.Lim, D.W.Potter, and J.Pawliszyn, **Envion. Lab.**, December/January 1992/1993, 4-10.

12. J.Ritter, V.K.Stromquist, H.T.Mayfield, M.V.Henley, and B.K.Lavine, "Solid Phase Microextraction for Monitoring Jet Fuel Components in Ground Water," **Microchem. J.**, 1995, submitted.

13. V.K.Stromquist, "Solid Phase Microextraction as a Method for Quantifying Jet Fuel Contamination in Water," Final Report for **Summer Graduate Research Program**, Armstrong Laboratory, 1994.

14. D.R.Burris, and W.G.MacIntyre, **Environ. Tox. Chem.**, 1985, 4, 371-377.

15. B.K.Lavine, P.Kromman, A.Faruque, H.T.Mayfield, "Source Identification of Fuel Spills by Pattern Recognition Analysis of High Speed Gas Chromatograms," **Anal. Chem.**, 67(20), IN PRESS.

Brother Lawless Report not available at time of publication.

A Statistical Analysis of Lead-Based Paint Survey Data

Tze-San Lee
Associate Professor
Department of Mathematics

Western Illinois University
Macomb, IL 61455

Final Report for:
Summer Faculty Research Program
Armstrong Laboratory

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base
Washington, D.C.

and

Armstrong Laboratory

August 1995

A STATISTICAL ANALYSIS OF LEAD-BASED PAINT SURVEY DATA

Tze-San Lee
Associate Professor
Department of Mathematics
Western Illinois University

Abstract

To identify and assess the lead-based paint (LBP), Galson corporation was retained by the Air Force to conduct the LBP survey. A stratified random sampling was adopted in the survey protocol. Prior to starting the survey, Calson's LBP Survey Manager, in consultation with Base Civil Engineering and Military Housing, reviewed information on construction and painting history for base housing. Based on this review, housing types was placed into different homogeneous exposure groups (HEGs). The Department of Housing and Urban Development (HUD) Interim Guideline was followed to determine the sample size. From each housing group, both the building and the location of painted components to be tested were randomly selected. The LBP survey was performed using SCIETEC Corporation's metal analysis probe X-ray fluoresence spectrum analyzer. A painted component was interpreted to contain lead if the testing result is greater than or equal to 1.3 mg Pb/cm^2, regardless what the duration of the testing was. From the 24 completed LBP surveys of the Air Combat Command bases, four were randomly selected for data analysis. This study indicates: (1) There is a marked variation in the prevalence rate of painted component containing lead in both the between-base and the within-base variation, (2) The HEGs assumption fails to hold for most of the housing groups, (3) The sample size determination used in the survey protocol is not optimal, (4) A systematic sampling with random start would work better than the simple random sampling in selecting the sampled building within the housing group, and (5) A ranking list is constructed for painted components likely to contain lead. This list can serve as a guide in any future LBP survey to select the painted components that are most likely to contain lead.

A STATISTICAL ANALYSIS OF LEAD-BASED PAINT SURVEY DATA

Tze-San Lee

Introduction

In response to increasing evidence that lead-based paint (LBP) presents a health risk to young children, the Department of Defense and the Air Force have issued policy and guidance on the management of facilities containing LBP. The policy establishes the need to prepare LBP management plans, perform risk assessments, and inspect, survey, and analyze high priority facilities for the presence containing lead. It is assumed that all facilities built before 1980 contain some amount containing lead which must be adequately managed. The Air Combat Command (ACC) action is to have all non-housing high priority facilities and a representative sample of all housing facilities surveyed for the presence containing lead. The ultimate goal of the project is to enable the Base Civil Engineer to properly program for future facility projects in addition to the beneficial effect of possible reduction of health risks for base personnel.

Galson Corporation was retained by the Air Force to identify and assess LBP. Out of the 27 ACC bases, the LBP survey works were completed for 24 bases in the United States and Panama. It is the purpose of this report to conduct a statistical analysis on the collected data to study the base-to-base variation of the LBP problem at the ACC bases.

Materials and Methods

Sampling Design

Basically, stratified random sampling was employed (Cochran [1]). Prior to starting the survey, Galson's LBP Survey Manager, in consultation with Base Civil Engineering and Military Housing, reviewed information on construction and painting history for base housing. Based on this review, housing types was placed into different LBP reporting groups, i.e., a stratification was employed. During the survey, Galson surveyors visually inspected housing units to verify the painted components to be tested and their qualities. Each different type of painted components of the selected building was tested and the sample locations for each component tested and their buildings were randomly selected. Some nonhousing

facilities, e.g., the Youth Center, the Child Development Center, the Child Care Center, and etc., were also tested. The HUD Guidelines was followed to determine the required sample size .

Sample Assays

The survey was conducted in accordance with the Department of Housing and Urban Development's (HUD) Lead-Based Paint Interim Guidelines for Hazard Identification and Abatement in Public and Indian Housing, revised May 1991. The LBP survey was performed using SCITEC Corporation's metal analysis probe (MAP) X-ray fluorescence (XRF) spectrum analyzers. Testing was performed by technicians trained and certified in the use of the MAP XRF.

Surveyors followed the testing procedures recommended by SCITEC for the MAP XRF. A summary of the three-step procedure is given as follows (Galson Corporation [3]):

Step 1. Conduct an initial 15-second "screen" assay at the sample location. This 15-second measurement should have an average precision (95 percent confidence level) of approximately ±0.6 mg Pb/cm^2 as determined by the National Bureau of Standards. The results of this 15-second screen samples were interpreted as follows :

Result	**Interpretation**
> 1.6 mg Pb/cm^2	Positive for lead
< 0.4 mg Pb/cm^2	Negative for lead
≥ 0.4 mg Pb/cm^2 and ≤ 1.6 mg Pb/cm^2	Inconclusive-Proceed to Step 2

Step 2. Conduct a 60-second "test" assay at the same location. The average precision for this assay is approximately ±0.3 mg Pb/cm^2. The results of this 60-second test were interpreted as follows:

Result	**Interpretation**
> 1.3 mg Pb/cm^2	Positive for lead
< 0.7 mg Pb/cm^2	Negative for lead
≥ 0.7 mg Pb/cm^2 and ≤ 1.3 mg Pb/cm^2	Inconclusive-Proceed to Step 3

Step 3. Conduct a 240-second "confirm" assay at the same location. The average precision for this assay is approximately ±0.15 mg Pb/cm^2. The results of this 240-second test were interpreted as follows:

Result	**Interpretation**
> 1.2 mg Pb/cm^2	Positive for lead
< 0.8 mg Pb/cm^2	Negative for lead
≥ 0.8 mg Pb/cm^2 and ≤ 1.2 mg Pb/cm^2	Inconclusive

HUD Guidelines define LBP levels between 0.8 and 1.2 mg Pb/cm^2 as being inconclusive and require that such samples be sampled and analyzed by a laboratory for lead content. However, bulk sampling was not included in the contract with Galson's Corporation.

Data Analysis

The data used in the analysis are provided by the Galson Corporation from their database "pcv3". Due to the time limitation, the author decides not to conduct a thorough analysis of all of the data collected at 24 ACC Bases. Instead, a two-stage cluster sampling is employed (Cochran [1]). At Stage 1, each of the 24 ACC Bases is viewed as a cluster, and then from the alphabetically ordered ACC Bases using a random digit table to select 4 bases. At Stage 2, a stratified random sampling was used. According to the order of selection they are listed as follows: Griffiss, Howard-Albrook, Sawyer, and Ellsworth. A conservative criterion is adopted to interpret a sample assay of the painted component to contain lead if the testing result is greater than or equal to 1.3 mg Pb/cm^2, regardless of what the sampling duration is. Let p_G, p_{HA}, p_S, and p_E be, respectively, the base-wide prevalence rate of painted components containing lead at Griffiss, Howard-Albrook, Sawyer, and Ellsworth Air Force Bases (AFBs). First, a natural question to ask is : are the prevalence rates the same for all four bases? This question can be re-phrased in terms of the statistical hypothesis testing, i.e., we wish to test the following null and alternative hypotheses:

$$H_0 : p_G = p_{HA} = p_S = p_E ,$$

versus

H_a : Not all of the prevalence rates are equal.

The chi-square test is applied to test the above hypotheses (Fleiss [2]). Second, in the sampling protocol all the military housings are assumed to belong to some homogeneous exposure groups (HEGs). We would like to use the collected data to verify if the assumption of HEGs holds. Again, the chi-square test is used

to test the homogeneity assumption by housing groups. Finally, the overall prevalence rates for painted components containing lead were calculated and ranked accordingly.

Results

We calculated the base-wide prevalence rates for four selected bases (Table 1). The rates ranged from the lowest 5.4% (Ellsworth AFB) to the highest 19.3% (Howard-Albrook AFB). Through the use of the chi-square test, the null hypothesis of equal rates is rejected at the significance level less than $\alpha= .001$. As far as the homogeneity assumption by housing group is concerned, only 4 housing groups, Hancock Field in Griffiss AFB, Albrook-New Tropical in Howard-Albrook AFB, Capehart (Increment 5) and Appropriated (Increment 6) in Sawyer AFB, are shown to satisfy the homogeneity assumption (Table 2).

The prevalence rates of painted components containing lead by housing and non-housing facilities were calculated (Table 3). Five unique housing units (#231, #232, #480, #880, and #883) in Griffiss AFB were grouped into a group called "Unique". For all four bases, the non-housing facilities such as the Youth Center, the Child Development Center, and the Child Care Center were grouped into a single group called "Base". By and large, the prevalence rate for the non-housing facilities is about the median of the housing groups for all four bases. As to the housing groups, only those of Ellsworth AFB are less than 10%. All of the other bases had some housing groups with the prevalence rates exceeding 10%. For example, all of the housing groups in Howard-Albrook AFB, three housing groups, Capehart (Increment 2, 2 Phase 4, and 3), in Sawyer AFB, and two housing groups, Hancock Field and Unique, in Griffiss AFB .

The prevalence rates by painted components for all four bases were calculated and given in columns 2-5 of Table 4 . Concerning the reliability of the estimated prevalence rate containing lead for painted components, the total number and number of positive lead-testing sample assays are also included in Table 4. A full description of the abbreviated name for painted components is given in Table 7. To interpret the prevalence rates by painted components, the sample size should also be taken into consideration. For instance, we should not put much faith in the prevalence rate (100%) for EDFO in Griffiss AFB and for BBOT in Ellsworth AFB because the sample size is so small, only one sample assay for both of them was collected. It is noted that the painted component with the highest prevalence rate varies among bases. To rank painted components across bases, the variation in their prevalence rates among bases has to be

accounted for. The across-base prevalence rates for those components whose sample assays were collected for at least two bases were calculated and given in the last column of Table 4. The ranking of the painted components with nonzero across-base prevalence rates is given in Table 5. However, please be reminded that the probability of correct ranking in Table 5 depends on both the difference between their prevalence rates and the sample size. For example, the probability of correct ranking between ETMT and CLOT is only .90 because with the difference between their prevalence rates being .48 the smallest sample size for the 95% correct ranking is 9 (Table E.1 with k=2 of Gibbons, et al [4]), but the sample size for ETMT is only 8.

Discussion

This study indicates:

•First, there is a substantial variation in the prevalence rate of painted components containing lead across bases. Even disregarding Howard-Albrook AFB located in Panama and considering only bases in the US. continent, the highest 13.1% (Sawyer AFB) is still 2.4 times the lowest 5.4% (Ellsworth AFB) . The reason for this variation is not clear. Marked variation in the prevalence rates also exists by housing groups within the base. Again, the reason is not clear.

•Second, note that the sample size determination given in the HUD Guidelines, based on the model of hypergeometric distribution, is a probability rather than statistical approach. Strictly speaking, it is inappropriate to be used in any sampling design. In fact, for stratified sampling, the optimum allocation is known to be proportionate allocation if the costs of collecting sample assays from painted components are assumed to be the same (Cochran [1]). However, except at Griffiss AFB in which approximately 6% of the buildings within each housing group were sampled, the sample sizes for different housing groups at the other three bases are very disproportionate (Table 6); hence clearly not optimal.

•Third, since most housing groups failed to pass the chi-square test for the homogeneity assumption, a natural question arises: can those non-homogeneous housing groups be further divided into homogeneous subgroups? To answer this question, the maps for each housing group were obtained. The prevalence rate for each sampled building was computed. Then each sampled building was marked, respectively, with three different colors, red, blue, and green, according to the prevalence rate "p": red for $p \geq 10\%$, blue for

$p \geq 5\%$, but $< 10\%$, and green for $p < 5\%$. After coloring for each sampled building was done, a further division into homogeneous subgroups for some housing groups is clearly revealed on the map. For example, at Ellsworth AFB, the Capehart Housing can be divided into three homogeneous subgroups: (1) Capehart 500 -, (2) Capehart 220 -, and (3) Capehart 190, -Housing Area ; the Nike Housing can be divided into two homogeneous subgroups: (1) East Nike -, and (2) South Nike and West Nike, -Housing Area; the Shell Housing could be divided into two homogeneous subgroups: (1) Buildings #5401, #6401, #6801, and (2) all the other buildings. A further division into homogeneous subgroups for non-homogeneous housing groups at the other bases is less clear from the current data shown on the housing map. Additional data would have helped our attempt.

•Fourth, we note that the buildings within the housing group are numbered systematically, the sampled building selected by a systematic sampling with single or multiple random start seems to work better than by simple random sampling.

•Finally, from Table 5 it is noted that some painted components with very low prevalence rates were collected with very large samples, e.g., WLSH (4765) and CLSH (3962). Using Table 5 as a guide, the sampling effort in future LBP surveys should concentrate on those painted components whose prevalence rates exceeded 11% .

Acknowledgments

This research was supported by a faculty summer research fellowship sponsored by the Air Force Office of Scientific Research and administered by Research & Development Laboratory. Fruitful discussions with Lt. Colonel C. Hollenbeck and Captain T. Mukoda, and that Galson Corporation and the Housing office at the selected bases kindly provides, respectively, the data and the housing maps are gratefully acknowledged. All the calculations was facilitated by using Microsoft Excel.

References

1. Cochran, W. G. (1977). Sampling Techniques, 3rd Edition, John Wiley & Sons, Inc., New York.

2. Fleiss, J.L. (1973). Statistical Methods for Rates and Proportions, John Wiley & Sons, Inc., New York.

3. Galson Corporation's Final Report (1995). Griffiss Air Force Base/ Lead-Based Paint Identification Survey, Department of the Air Force, F33615-89-4001, Order 0036, Syracuse, New York.

4. Gibbons, J. D., Olkin, I., and Sobel, M. (1977). Selecting and Ordering Populations: A New Statistical Methodology, John Wiley & Sons, Inc., New York.

5. US. Department of Housing and Urban Development (1990). Lead-based paint: Interim guidelines for hazard identification and abatement in public and Indian housing, Washington, DC..

Air Force Base	# of positive lead-testing sample assays	# of negative lead-testing sample assays	% of lead-containing sample assays
Griffiss	525	5010	9.5
Howard-Albrook	1976	8270	19.3
Sawyer	2117	14085	13.1
Ellsworth	696	12143	5.4

Table 1 The prevalence rate by base.

Air Force Base	Housing group	Chi-square test (degrees of freedom)	Is the test significant at $\alpha = .05$?
Griffiss	Hancock Field	18.3 (12)	No
	Skyline Terrace	83.4 (26)	Yes
	Woodhaven Village	51.0 (15)	Yes
Howard-Albrook	Albrook-New Tropical	8.8 (11)	No
	Albrook-Tropical	171.1 (22)	Yes
	Howard-Gateway	65.2 (25)	Yes
	Howard-Tropical	126.9 (31)	Yes
Sawyer	Capehart (Increment 1)	76.2 (30)	Yes
	Capehart (Increment 2)	125.1 (13)	Yes
	Capehart (Increment 2 Phase 4)	52.1 (13)	Yes
	Capehart (Increment 3)	133.0 (35)	Yes
	Capehart (Increment 4)	108.1 (32)	Yes
	Capehart (Increment 5)	20.0 (13)	No
	Appropriated (Increment 6)	8.6 (9)	No
Ellsworth	Capehart	289.4 (57)	Yes
	Nike	71.5 (37)	Yes
	Shell	210.8 (44)	Yes

Table 2 A test on the homogeneity assumption by housing group.

(a) Griffiss Air Force Base

Housing and nonhousing group	# of positive lead-testing sample assays	# of negative lead-testing sample assays	% of lead-containing sample assays
Hancock Field	129	1083	10.6
Skyline Terrace	52	1831	2.8
Woodhaven Village	92	1044	8.1
Unique	208	549	27.5
Base	43	455	8.6

(b) Howard-Albrook Air Force Base

Housing and nonhousing group	# of positive lead-testing sample assays	# of negative lead-testing sample assays	% of lead-containing sample assays
Albrook-Tropical	682	1547	30.6
Albrook-New Tropical	138	898	13.3
Howard-Gateway	575	2600	18.1
Howard-Tropical	424	2560	14.2
Base	157	645	19.6

(c) Sawyer Air Force Base

Housing and nonhousing group	# of positive lead-testing sample assays	# of negative lead-testing sample assays	% of lead-containing sample assays
Capehart (Increment 1)	303	3175	8.7
Capehart (Increment 2)	242	1260	16.1
Capehart (Increment 2 Phase 4)	313	1164	21.2
Capehart (Increment 3)	1082	2932	27.0
Capehart (Increment 4)	94	3114	2.9
Capehart (Increment 5)	17	1210	1.4
Appropriated (Increment 6)	1	715	0.1
Base	65	519	11.1

(d) Ellsworth Air Force Base

Housing and nonhousing group	# of positive lead-testing sample assays	# of negative lead-testing sample assays	% of lead-containing sample assays
Capehart	246	4300	5.4
Nike	201	2611	7.2
Shell	231	4978	4.4
Base	18	274	6.2

Table 3 The prevalence rate by housing or nonhousing group.

Component	Percent (# of components containing lead: total # of sampled components)				
	Griffiss	How-Alb	Sawyer	Ellsworth	Across-base
BBOT	0 (0:1)	33 (3:9)		100 (1:1)	36 (4:11)
BBWD	11 (58:534)	20 (66:333)	37 (417:1134)	2 (26:1291)	17 (567:3292)
BCMT		10 (3:30)	62 (77:125)	60 (62:103)	55 (142:258)
BCWD		67 (2:3)	23 (15:66)	0 (0:10)	22 (17:79)
CBBMT		0 (0:22)			
CBMT		20 (33:165)	0 (0:4)		20 (33:169)
CBWD	8 (7:87)	35 (77:219)	16 (15:91)	0.3 (1:353)	13 (100:750)

CBWV		0 (0:58)	0 (0:17)		0 (0:75)
CDMT		6 (1:16)	0 (0:1)	0 (0:5)	4.5 (1:22)
CDPWD			15 (49:326)		
CDWD	4 (8:202)	40 (10:25)	1 (2:181)	0 (0:78)	4.2 (20:486)
CDWV			0 (0:127)		
CLASP		15 (35:230)			
CLCT		17 (64:378)			
CLMD		40 (84:210)			
CLMT		0 (0:5)			
CLOT	20 (2:10)	0 (0:5)		53 (31:58)	45.2 (33:73)
CLPL	3 (8:262)			0 (0:6)	3 (8:268)
CLSH	3 (11:376)	1 (3:331)	0.1 (2:1695)	1 (19:1560)	0.9 (35:3962)
CLTX			0 (0:2)		
CLWD	39 (7:18)	17 (51:306)	0 (0:4)	10 (3:29)	17.1 (61:357)
DFMT	6 (3:49)	4 (2:46)	11 (5:47)	2 (1:64)	5.3 (11:206)
DFWD	9 (53:581)	27 (168:625)	23 (156:672)	2 (38:1583)	12 (415:3461)
DJMT	20 (37:188)	0 (0:5)	50 (3:6)	0 (0:3)	19.8 (40:202)
DJWD	6 (7:120)	27 (62:233)	25 (91:362)	15 (4:26)	22.1 (164:741)
DRAP	0 (0:33)				
DREIM	10 (8:77)	3 (3:91)	0.5 (1:188)	11 (21:184)	6.1 (33:540)
DREIW		11 (45:421)	25 (75:301)	55 (56:101)	21.4 (176:823)
DREJM		17 (3:18)	0 (0:3)	0 (0:1)	13.6 (3:22)
DREJW		42 (75:178)	43 (183:421)	23 (80:354)	35.5 (338:953)
DREOM	5 (1:22)	10 (11:131)	0 (0:181)	12 (23:198)	6.6 (35:532)
DREOW	19 (15:77)	26 (114:436)	40 (134:332)	36 (72:198)	32.4 (335:1033)
DRIM	25 (7:28)		2 (4:188)	0 (0:8)	4.9 (11:224)
DRIW	31 (82:262)	20 (244:1196)	14 (197:1412)	8 (24:298)	32 (547:1710)
DRWV		1 (1:135)	0 (0:185)		0.3 (1:320)
EDFM	0 (0:2)	29 (6:21)	60 (3:5)	0 (0:2)	30 (9:30)
EDFO	100 (1:1)				
EDFW	63 (17:27)	52 (13:25)	44 (44:101)	29 (12:42)	44.1 (86:195)
EDRCT		15 (3:20)			
EDRWT	11 (5:41)			45 (23:51)	29.5 (28:95)
EDUCT		0 (0:16)	16 (14:87)		13.6 (14:103)
EFBL	0 (0:3)	6 (5:88)	0 (0:3)	0 (0:17)	4.5 (5:111)
EFCB		0 (0:1)			
EFCM		0 (0:1)			
EFCT		12 (16:132)			
EFCW	9 (2:23)	25 (1:4)	0 (0:24)	0 (0:3)	5.6 (3:54)
EFOT	0 (0:2)		0 (0:1)		
EFWD	46 (22:48)	17 (1:6)	39 (44:112)	11 (6:53)	33.3 (73:219)
EMISC		16 (13:80)	0 (0:4)	0 (0:26)	11.8 (13:110)
ESFC		11 (5:46)			
ESFO		100 (3:3)			
ESFW	55 (18:33)	40 (21:52)	100 (23:23)	63 (33:52)	59.4 (95:160)
ESTC		18 (5:28)			
ETMT		0 (0:3)	80 (4:5)		50 (4:8)
ETOT	0 (0:4)				
ETWD	56 (20:36)	34 (21:62)	70 (74:106)	43 (29:68)	52.9 (144:272)

EWSB		11 (14:127)			
EWWD	31 (5:16)	61 (11:18)	86 (25:29)	4 (2:47)	39.1 (43:110)
FLCC	2 (1:50)	0 (0:3)	5 (5:111)	6 (5:81)	4.5 (11:245)
FLWD	0 (0:5)		0 (0:2)		
FLWV			0.1 (1:885)		
IDUCT		0 (0:25)	0 (0:2)		0 (0:27)
IMISC		0 (0:14)	0 (0:3)		0 (0:17)
IPWD		0 (0:2)			
PMISC		26 (15:57)			
RAMT		0 (0:6)	0 (0:17)	3 (2:68)	2.2 (2:91)
SHMT		0 (0:2)	0 (0:1)		
SHWD	3 (9:353)	13 (77:583)	12 (121:1049)	0.6 (6:1056)	7 (213:3031)
SSWD	6 (13:233)	16 (70:439)	14 (121:865)	0.6 (4:712)	9.7 (208:2149)
STCTZ		14 (8:56)			
STMTR		100 (3:3)			
STWBB		3 (1:33)	4 (6:139)		4.1 (7:172)
STWDB		100 (2:2)			
STWDP		0 (0:29)			
STWDR		27 (12:44)	6 (16:273)	6 (9:147)	8 (37:464)
STWDS		4 (1:23)	10 (26:264)	32 (11:34)	11.8 (38:321)
STWDT	0 (0:19)	6 (2:33)	6 (17:288)	14 (21:155)	8.1 (40:495)
STWDZ	2 (1:45)	0 (0:32)	8 (23:300)	6 (13:228)	6.1 (37:605)
WCRW		44 (14:32)	0 (0:2)		41.2 (14:34)
WFMT		0 (0:1)			
WFOT				0 (0:13)	
WFWD	10 (53:549)	18 (24:133)	6 (63:1020)	1 (18:1241)	5.4 (158:2943)
WHWD	0 (0:1)	7 (5:71)			6.9 (5:72)
WJWD	29 (2:7)	12 (5:41)	6 (1:17)		12.3 (8:65)
WLBL	8 (5:59)	11 (20:182)	0 (0:36)	4 (6:166)	7 (31:443)
WLCT		28 (303:1066)	8 (16:198)		25.2 (319:1264)
WLMT		0 (0:2)			
WLOT	12 (6:52)	7 (1:14)	0 (0:35)	3 (3:95)	4.6 (9:196)
WLPA	5 (1:21)	13 (2:16)	10 (3:31)	3 (1:40)	6.5 (7:108)
WLPL	2 (5:284)			0 (0:8)	1.7 (5:292)
WLSH	2 (12:526)	1 (6:435)	0.2 (4:1949)	2 (28:1855)	1 (50:4765)
WLWD	10 (3:31)	12 (12:101)	15 (12:80)	0 (0:30)	11.2 (27:242)
WSASH		14 (2:14)			
WSWD	13 (7:55)	52 (93:178)	39 (25:64)	11 (2:19)	40.2 (127:316)

Table 4 The prevalence rate by painted component.

Painted component	Total # of collected sample assays	% of lead-containing sample assays
ESFW	160	59.4
BCMT	258	55.0
ETWD	272	52.9
ETMT	8	50.0

CLOT	73	45.2
EDFW	195	44.1
WCRW	34	41.2
WSWD	316	40.2
EWWD	110	39.1
BBOT	11	36.4
DREJW	953	35.5
EFWD	219	33.3
DREOW	1033	32.4
DRIW	1710	32.0
EDFM	30	30
EDRWT	95	29.5
WLCT	1264	25.2
DJWD	741	22.1
BCWD	79	21.5
DREIW	823	21.4
DJMT	202	19.8
CBMT	169	19.5
BBWD	3292	17.2
CLWD	357	17.1
DREJM	22	13.6
EDUCT	103	13.6
CBWD	750	13.3
WJWD	65	12.3
DFWD	3461	12.0
STWDS	321	11.8
EMISC	110	11.8
WLWD	242	11.2
SSWD	2149	9.7
STWDT	495	8.1
STWDR	464	8.0
SHWD	3031	7.0
WLBL	443	7.0
WHWD	72	6.9
DREOM	532	6.6
WLPA	108	6.5
STWDZ	605	6.1
DREIM	540	6.1
EFCW	54	5.6
WFWD	2943	5.4
DFMT	206	5.3
DRIM	224	4.9
WLOT	196	4.6
CDMT	22	4.5
EFBL	111	4.5
FLCC	245	4.5
CDWD	486	4.2
STWBB	172	4.1
CLPL	268	3.0

RAMT	91	2.2
WLPL	292	1.7
WLSH	4765	1.0
CLSH	3962	0.9
DRWV	320	0.3

Table 5 The across-base ranking of painted components with nonzero prevalence rate.

Air Force Base	Housing group	No. of sampled buildings (No. of total buildings)	% of sampled buildings
Griffiss	Hancock Field	13 (216)	6.0
	Skyline Terrace	27 (460)	5.9
	Woodhaven Village	16 (270)	5.9
Howard-Albrook	Albrook-New Tropical	12 (50)	24
	Albrook-Tropical	24 (449)	5.3
	Howard-Gateway	26 (243)	10.7
	Howard-Tropical	32 (430)	7.4
Sawyer	Capehart (Increment 1)	31 (323)	9.6
	Capehart (Increment 2)	14 (101)	13.9
	Capehart (Increment 2 Phase 4)	14 (111)	12.6
	Capehart (Increment 3)	37 (342)	10.8
	Capehart (Increment 4)	34 (368)	9.2
	Capehart (Increment 5)	15 (100)	15
	Appropriated (Increment 6)	20 (200)	10
Ellsworth	Capehart	58 (911)	6.4
	Nike	38 (48)	79.2
	Shell	45 (100)	45

Table 6 Percentage of sampled buildings by housing group.

Component	Description
BBOT	BASEBOARDS/OTHER
BEWD	Baseboards/Wood
BCMT	BEAM OR COLUMN/METAL
BCWD	BEAM OR COLUMN/WOOD
CBBMT	CIRCUIT BREAKER BOX COVER/METAL
CBMT	Cabinet Door/Metal
CBOT	Cabinet Door/Other
CBWD	Cabinet Door/Wood
CBWV	VARNISHED CABINETS/WOOD
CDMT	Closet Door/Metal
CDOT	Closet Door/Other
CDPWD	Closet Door/Pressed Wood
CDWD	Closet Door/Wood
CDWV	VARNISHED CLOSET DOOR/WOOD
CHPM	CLOSET HANGER POST/METAL
CHPW	CLOSET HANGER POST/WOOD
CLASP	CEILING/FIBERBOARD
CLCT	CEILING/CONCRETE
CLMD	CEILING MOLDING
CLMT	CEILING METAL
CLOT	Ceiling/Other
CLPL	Ceiling/Plaster
CLSH	Ceiling/Sheetrock
CLTX	CEILING/TEXTURED
CLWD	CEILING/WOOD
DFMT	Door Frame/Metal
DFOT	Door Frame/Other
DFWD	Door Frame/Wood
DJMT	DOOR JAMB/METAL (INTERIOR)
DJOT	DOOR JAMB/OTHER (INTERIOR)
DJWD	DOOR JAMB/WOOD(INTERIOR)
DRAP	ACCESS PANEL DOOR/WOOD
DREIM	Door Exterior (Inside) /Metal
DREIW	Door Exterior (Inside)/Wood
DREJM	EXTERIOR DOOR JAMB/METAL
DREJW	EXTERIOR DOOR JAMB/WOOD
DREOM	Door Exterior (Outside)/Metal
DREOW	Door Exterior (Outside)/Wood
DRIM	Door Interior/Metal
DRIO	Door Interior/Other
DRIW	Door Interior/Wood
DRWV	VARNISHED DOOR/WOOD
EBCT	Exterior Basement Foundation/Concrete
EDFM	Exterior Door Frame/Metal
EDFO	Exterior Door Frame/Other
EDFW	Exterior Door Frame/Wood
EDRCT	THRESHOLD/CONCRETE
EDROT	THRESHOLD/OTHER
EDRWT	Threshold/Wood
EDUCT	EXTERIOR DUCT
EFBL	EXTERIOR FACADE/BLOCK OR BRICK
EFCB	EXTERIOR FENCING/BLOCK
EFCM	EXTERIOR FENCE/METAL
EFCPW	EXTERIOR FENCE/PRESSED WOOD
EFCT	EXTERIOR FACADE/CONCRETE
EFCW	Exterior Fence/Wood
EFMT	EXTERIOR FACADE/METAL
EFOT	Exterior Facade/Other
EFSH	Exterior Facade/Shingle
EFTR	EXTERIOR FACADE/TRANSITE
EFWD	Exterior Facade/Wood
EGUT	EXTERIOR GUTTER
EMISC	MISC/EXTERIOR
EPMT	Exterior Post/Metal
ERMT	EXTERIOR RAILING/METAL
ERWD	EXTERIOR RAILING/WOOD
ESDIM	EXTERIOR STORM DOOR/METAL (INSIDE)
ESDOM	EXTERIOR STORM DOOR/METAL (EXTERIOR)
ESFC	EXTERIOR SOFFITTS/CONCRETE
ESFM	EXTERIOR SOFFITTS/METAL
ESFO	Exterior Soffitts/Other
ESFS	Exterior Soffitt/Sheetrock
ESFT	EXTERIOR SOFFITS/TRANSITE
ESFW	Exterior Soffitts/Wood
ESTC	EXTERIOR STEPS/CONCRETE
ESTW	EXTERIOR STEPS WOOD
ETMT	EXTERIOR TRIM/METAL
ETOT	Exterior Trim/Other
ETWD	Exterior Trim/Wood
EWMT	Exterior Window Frame/Metal
EWOT	Exterior Window Frame/Other
EWSB	EXTERIOR WINDOW SILL/BLOCK
EWWD	Exterior Window Frame/Wood
FLCC	Floor/Concrete
FLWD	Floor/Wood
FLWV	VARNISHED FLOOR/WOOD
FRAP	ACCESS PANEL DOOR/WOOD
IDUCT	INTERIOR DUCT OR PIPE
IMISC	MISC/INTERIOR
IPMT	INTERIOR POST/METAL
IPWD	INTERIOR POST/WOOD

LINT	LINTEL/METAL
MANT	FIREPLACE MANTEL/WOOD
PLMB	METAL BENCHES
PLMP	METAL STEPS
PLMS	METAL STRUCTURE
PLMT	METAL TRIM
PLPM	METAL PICNIC TABLE
PLPO	OTHER PICNIC TABLE
PLPW	WOODEN PICNIC TABLE
PLRM	METAL RIDERS
PLRO	OTHER RIDERS
PLWB	WOODEN BENCHES
PLWP	WOODEN STEPS
PLWS	WOODEN STRUCTURE
PLWT	WOODEN TRIM
PMISC	MISCELLANEOUS PLAYGROUND EQUIPMENT
RAMT	RADIATOR/METAL
SHMT	Shelf/Metal
SHUW	SHUTTERS/WOOD
SHWD	Shelf/Wood
SSMT	Shelf Support/Metal
SSWD	Shelf Support/Wood
STCTZ	STAIR RISER/CONCRETE
STMTR	STAIR RAIL/METAL
STWBB	STAIR BASEBOARD/WOOD
STWDB	Stair Baluster/Wood
STWDP	Stair Post/Wood
STWDR	Stair Railing/Wood

STWDS	Stair Stringer/Wood
STWDT	Stair Tread/Wood
STWDZ	Stair Riser/Wood
VENT	AIR VENT/METAL
WCHW	WINDOW CURTAIN HOLDERS/WOOD
WCRW	WALL CHAIR RAIL/WOOD
WFMT	Window Frame/Metal
WFOT	Window Frame/Other
WFWD	Window Frame/Wood
WHWD	Window Header/Wood
WJMT	Window Jamb/Metal
WJWD	Window Jamb/Wood
WLBL	Wall/Block
WLCT	WALL/CONCRETE
WLMT	WALL/METAL
WLOT	Wall/Other
WLPA	Wallpaper
WLPL	Wall/Plaster
WLSH	Wall/Sheetrock
WLTR	WALL/TRANSITE
WLWD	Wall/Wood or Paneled
WSASH	WINDOW SASH
WSMT	Window Sill/Metal
WSOT	Window Sill/Other
WSWD	Window Sill/Wood
WWEL	WINDOW WELL/WOOD

Table 7 A description for the code name of painted components.

EVALUATING ENVIRONMENTAL SYSTEM: CORRECTING FOR THERMAL GRADIENT IN TIME ORDERED REPEATED MEASURES DESIGNS

David A. Ludwig
Associate Professor
Department of Mathematical Sciences

University of North Carolina, Greensboro
1000 Spring Garden Street
Greensboro, NC 27412-5001

Final Report for:
Summer Faculty Research Program
Armstrong Laboratory

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, DC

and

Armstrong Laboratory

August 1995

EVALUATING ENVIRONMENTAL SYSTEM: CORRECTING FOR THERMAL GRADIENT IN TIME ORDERED REPEATED MEASURES DESIGNS

David A. Ludwig
Associate Professor
Department of Mathematical Sciences
University of North Carolina Greensboro

Abstract

Extension of several difference score approaches for the analysis of time ordered repeated measures experiments involving more than two time periods are presented and discussed. Experiments of this type are commonly encountered in the study of environmental stresses such as heat and cold. The proposed method incorporates statistical correction for the problem of thermal gradient, and is easily understood and interpreted by research workers. It is also mathematically equivalent to more complex and often confusing split-plot repeated measures formats.

EVALUATING ENVIRONMENTAL SYSTEM: CORRECTING FOR THERMAL GRADIENT IN TIME ORDERED REPEATED MEASURES DESIGNS

David A. Ludwig

INTRODUCTION

Laird (1983), proposed an alternative approach to the analysis of pre-post, two or more group, repeated measures designs. The suggested method, uses the difference score calculated between the two time periods in which the experimental unit was repetitively measured. The difference score is used as the dependent variable with the pretest measure acting as a covariate. Using the pretest measure as a covariate, equates treatment groups at their starting point and, according to Laird (1983), corrects for regression toward the mean.

Regression toward the mean results when the dependent measure is subject to measurement error, and the observed values on the pretest measure tend to be higher or lower than the pretest true scores. Subsequent measures in time will appear to rise or fall depending on the degree of over or under estimation at the pretest (Cohen and Cohen 1983). Although Laird (1983), specifically addresses the consequences of regression toward the mean, her procedure can be used to ~~corrects~~ correct for differences in pretest values arising from other causes.

Since a difference score is used as the dependent measure, adjusted group means can be tested that each is equal to zero by

placing the group means over their respective standard errors. Or, confidence intervals can be constructed around the group means providing a range of possible outcome values. The between groups test, addresses the question of interaction (are the pre-posttest differences different for each treatment group), while the test that each group's adjusted mean is equal to zero can be used to evaluate within group changes. In practice, this approach is easy to explain, program, and interpret, while being intuitively appealing. This is especially helpful when working with researchers with limited and varied statistical backgrounds.

Alternative Models

Traditionally, a split-plot format (group and time main effects and subsequent interaction) would be used to analyze such an experiment (Brogan and Kutner 1980). Although the split-plot approach and the covariance difference score approach (Laird 1983) are both correct, the standard split-plot approach (without a covariate) does not correct for differences at the pretest, and requires the interpretation of a two-factor interaction (Huck and McLean 1975). A constant covariate (constant at each repeated measurement) such as a pretest measurement can be added to the split-plot model, at the expense of much greater complexity. The basic difference score approach (Laird's method without the pretest covariate) reduces the complexity of the split-plot statistical model while retaining information regarding within group changes. Differencing the two repeated measures eliminates

the two factor interaction and incorporates this information into the test of between group effects.

The covariance difference score approach (Laird 1983), also seems to be an improvement over the standard analysis of covariance model. The standard ANCOVA approach uses the posttest as the dependent variable while the pretest is used as the covariate. It corrects for pretest values, but does not directly estimate the degree of change within each group (Bock 1975). It is more of a classical analysis of covariance approach than a repeated measures analysis. This approach might be preferred when the investigator is only interested in posttest differences. Researchers tend to dislike this approach, since testing for within group change over the repeated measurements is of primary interest. Although the resulting test statistic for the group effect in the standard ANCOVA approach will be equal to that obtained on the group effect using the covariance difference score approach, researchers often complain that information on within group change is lost, since they do not understand how this information is incorporated into the standard ANCOVA model.

Extending the Pretest-Posttest Model

Since the covariance difference score approaches is intuitively appealing and easily explained to research workers, an extension of this technique to situations in which the experimental units are measured across more than two time periods would be useful. Repetitive differencing would not provide an extension to these

methods, since the crossed effect (i.e., repeated measures) would only be reduced by one level and the split-plot format would still need to be used. The problems associated with the assumption of an equal variance equal correlation pattern of the covariance matrix would also need to be addressed (Morrison 1990). This in turn, may lead to a multivariate approach that is difficult to interpret for most researchers and in many cases not defined. Brogan and Kutner (1980) state, "There is no logical extension of the basic difference score approach when there are more than two levels of the repeated measures." Fortunately, this is not exactly true.

When the repeated measures are time ordered, a linear function can be fit across the time periods separately for each subject, and the resulting slope estimates used as the measure of gain. Slope calculations do not require any assumptions regarding the variance covariance pattern of the repeated measures. All that is required is that each experimental unit be observed at the same points in time. In the event a linear function is not a good representation of the gain, higher order functions based on polynomial regression can be used. This procedure is synonymous with polynomial decomposition used within the split-plot repeated measures approach (Winer 1971). A one-way analysis of variance on the slopes using treatment group as the independent variable, is mathematically equivalent to orthogonal polynomial partitioning of the group by time interaction in the split-plot analysis and testing the linear component of the interaction with

the linear part of the within subjects error term. This is the preferred approach when time is a factor in the design, since some form of a linear or higher order polynomial function would be expected if true variation across time is being observed (Dawkins 1983). Furthermore, linear or higher order partitioning of the group by time interaction and the within subjects error term is required before valid statistical tests can be performed (Winer 1971). As with the two period difference model, mean slope estimates within a treatment group can be tested against zero or confidence limits can be determined.

A Flaw in Laird's Idea

There is a major problem with the approach suggested by Laird (1983). Since regression toward the mean is the result of measurement errors in the dependent variable, by definition, using it as an independent effect produces biases when estimating treatment means (Snedecor and Cochran 1989). Although the consequences of this action are well beyond the scope of this paper, the fact that the covariate was chosen because of its less than perfect reliability cannot be ignored. Statistical correction cannot account in any systematic way for regression toward the mean. Since regression toward the mean is a result of random measurement errors associated with the pretest, there is no way to tell if subsequent correction of posttest scores by pretest values are correct. Since posttest measures are being corrected using fallible pretest scores, adjusted posttest measures are just as fallible as the pretest. The technique

suggested by Laird (1983) could be used to adjust for differences in the pretest measure arising from more direct systematic influences (e.g., inequalities associated with intact groups in observational studies), but it does not adjust for regression toward the mean. Pedhazur and Schmelkin (1991), give an excellent account of why, using the pretest for covariate adjustment, does not correct the problem of regression toward the mean.

Evaluating Environmental Systems

Unlike regression toward the mean, which is purely a measurement error artifact, there are other, more physical reasons why responses over time might increase or decrease solely as a consequence of the starting value. Researchers in environmental science must contend with such situations when studying the effects of different environmental systems on temperature regulation in humans. Studies which evaluate the heat flow characteristics of different clothing materials, air flow rates, and work loads are all examples of typical problems studied by researchers interested in adaptation to environmental stress.

In one form or another, heat flow (how fast and/or how much), is of primary concern when evaluating environmental systems. Heat flow is proportional to the temperature gradient and to the thermal conductance of the medium through which the heat is passing (Richards 1973). Since the medium through which heat is passing (e.g., clothing, skin, air, etc.) is generally what is

being evaluated, it is important that the temperature (thermal) gradient be controlled. The thermal gradient is synonymous with the electrical potential gradient. Just as electrical current increases as electrical potential increases, heat loss or gain (e.g., flow) increases with increasing temperature difference between the environment and the organism. It is therefore important when testing systems that are designed to inhibit or augment heat flow, to control for the initial temperature of the organism.

In practice, initial temperature of warm blooded organisms is difficult to control. Individual differences, anxiety, and differential lag time between instrumentation and testing create substantial variation in initial temperature. Since design efforts to control initial temperature often fail, some type of statistical correction is usually required.

One common method encountered in the literature is to take a baseline measure of initial temperature and reference all subsequent measures over time to this baseline value. Often expressed as percentage change from baseline, this approach has several drawbacks, including an additive to multiplicative scale change, and lost information concerning the original metric (e.g., degrees). Laird's approach on the other hand seems well suited for this type of problem. Unlike correction for regression toward the mean, the covariate (baseline) is chosen for its known physical effect and is not included as a

consequence of measurement fallibility. Of course there still may be a measurement problem, but it is not automatic as it is when attempting to solve the problem of regression toward the mean. Change across more than two periods of time can easily be quantified with slope estimates calculated from the responses of each experimental unit.

SUMMARY

Although presented in the context of the thermal gradient, the proposed approach can be used in most any situation in which the researcher feels the need to compare responses over time adjusted for baseline values. Baseline values may be obtained prior to the experimental conditions or during the initial equilibrium phase.

Space will not permit an illustration of all the equivalent split-plot F-tests that can be generated through the analysis of least squares estimates, sums, or differences. Equivalent split-plot main effects tests have been illustrated by Brogan and Kutner (1980). While Laird (1983), presents an approach for the analysis of adjusted treatment means. The time by treatment group interaction is, in general, why an experiment is performed. For repeated measures experiments, the abilitity to detect differential gain or decline over time, is the primary concern. The question of interaction must also be addressed prior to the evaluation of main effects. Fortunately, relatively complex ANOVA can be simplified without any loss of information.

Complex split-plot repeated measures layouts are well beyond the training of most research workers, while the concepts of slope and the one-way ANOVA are more readily understood. Most research workers can obtain slopes, and run a one-way ANOVA on these slopes without much assistance. A covariance model (pretest or baseline adjustment) does add some complexity, but it is a relatively minor conceptual jump from the basic one-way ANOVA. The added complexity of adding a covariate to the statistical model is beyond the capabilities of most consulting statisticians, let alone the typical research worker.

More often than not, the worth and utility of a statistical procedure are evaluated on a purely mathematical bases, with little or no concern given to issues such as interpretability, intuitive appeal, ease of programming, or flexibility. The statistical methods advocated here are easily understood and can be performed by a majority of research workers. Much more so than split-plot approaches involving multiple error terms, interactions, and orthogonal polynomials. Or, multivariate approaches which are usually not fully understood by the typical researcher.

REFERENCES

Bock, D. R. (1975), *Multivariate Statistical Methods in Behavioral Research*, New York: McGraw-Hill, 493-495.

Brogan, D. R., and Kutner, M. H. (1980), "Comparative Analyses of Pretest-Posttest Designs," *The American Statistician*, 34, 229-232.

Cohen, J., and Cohen, P. (1983), *Applied Multiple Regression Correlation Analysis for the Behavioral Sciences*, (2nd ed.), New Jersey: Erlbaum, 45-46.

Dawkins, H.C. (1983), "Multiple Comparisons Misused: Why so Frequently in Response-Curve Studies," *Biometrics*, 39, 789-790.

Huck, S.W., and McLean, R.A. (1975), "Using a Repeated Measures ANOVA to Analyze the Data from a Pretest-Posttest Design: A Potentially Confusing Task," *Psychological Bulletin*, 82, 511-518.

Laird, N. (1983), "Further Comparative Analyses of Pretest-Posttest Research Designs," *The American Statistician*, 37, 329-330.

Pedhazur, E.J., and Schmelkin, L.P. (1991), *Mesaurement Design and Analysis An Integrated Approach*, New Jersey: Erlbaum, 284-286.

Richards, S.A. (1983), *Temperature Regulation*, London: Wykeham, 26-27.

Roberts, J.S., and Laughlin, J.E. (1983), Analyzing Analysis of Covariance Models in Univariate Repeated Measures Designs Using SAS, *Proceeding of the Annual SAS Users Group International Conference*, 8, 640-643.

Snedecor, G.W., and Cochran, W.G. (1989), *Statistical Methods,* (8th ed.), Ames: Iowa State Univ. Press, 173-174.

Winer, B.J. (1971), *Statistical Principles in Experimental Design*, (2nd ed.), New York: McGraw-Hill, 577-594.